Synthesis and Characterization of BN-tryptophan and its Incorporation into Proteins & the Cation-π Binding Ability of BN-indole:

Author: Katherine Lynn Michelle Boknevitz

Boston College

Morrissey Graduate School of Arts and Sciences

Department of Chemistry

SYNTHESIS AND CHARACTERIZATION OF BN-TRYPTOPHAN AND ITS INCORPORATION INTO PROTEINS & THE CATION-π BINDING ABILITY OF BN-INDOLE

A dissertation

By

Katherine Lynn Michelle Boknevitz

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

January 2020

SYNTHESIS AND CHARACTERIZATION OF BN-TRYPTOPHAN AND ITS INCORPORATION INTO PROTEINS &

THE CATION-π BINDING ABILITY OF BN-INDOLE

By

KATHERINE LYNN MICHELLE BOKNEVITZ

Dissertation Advisor:

Professor Shih-Yuan Liu

ABSTRACT: Described herein are two projects on the application and effects of BN/CC isosterism on indole-containing compounds. In the first chapter, the synthetic route to an unnatural boron and nitrogen-containing analogue of tryptophan (BN-tryptophan) via late-stage functionalization of BN-indole is disclosed and its spectroscopic properties are reported with respect to the natural amino acid, tryptophan. The incorporation of BN-tryptophan into proteins expressed in *E. coli* using selective pressure incorporation, a residue specific method of unnatural amino acid incorporation, is then reported and its reactivity and fluorescence in the proteins characterized. In the second chapter, the synthesis of a BN-indole-containing aromatic scaffold is reported and the cation-π binding ability characterized by nuclear magnetic resonance (NMR) monitored titrations is disclosed. The resulting chemical shifts were analyzed using a non-linear curve fitting procedure and the extracted association constants (K_a's) compared with the natural indole scaffold. Computations were also performed to support the titration results.

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Professor Shih-Yuan Liu, for his guidance and encouragement throughout my graduate school journey. I would also like to thank Professor Abhishek Chatterjee for his guidance in our collaborative BN-tryptophan work as well as for taking the time to serve on my thesis committee. Thanks to Professor James Morken for taking the time to serve on my thesis committee as well and for his invaluable insights and advice over the years.

Thank you to Dr. Andrew Baggett, Dr. Jacob Ishibashi, and Dr. Alec Brown who welcomed me to the Liu lab and guided me in the beginning of my time in the Liu lab. It was great having such helpful senior colleagues who were willing to answer all of my questions. Thanks to the Biomedical subgroup, especially Dr. Hyelee Lee and Dr. Peng Zhao, for their input on my project and assistance in lab.

I didn't only have great colleagues to work with in lab, but have made great friends who made such a challenging program more enjoyable. I greatly appreciate the camaraderie and friendship of my classmate, Dr. Cameron McConnell, as well as the younger members of the lab, Ricky Alvarado, Jon Deegan, Sierra Bentley, and the rest of the Liu lab crew, past and present. Special thanks to Jeremy Armand for pushing me to the finish line in these final months and being an overall amazing friend. Thanks to Kevin Byrne, for reading my thesis and his input on my work. Shout-out to Dr. Kelly McCarthy as well, whom I met through grad school, for being a great and understanding roommate and friend.

I would also like to thank Dale Mahoney, Lynne Pflaumer, Dr. Ian Parr, Dr. Marik Domin, and the rest of the BC office and instrumentation staff for keeping everything running smoothly. Thanks in particular to Dr. Thusitha Jayasundera (TJ) in the NMR lab for his help in the last few months with my NMR experiments.

My family played an integral role in this process. I could not have done this without the love and support of my wonderful parents, Laura and Pat, and my brother, Brian. Their perseverance through the many challenges they have faced over the years is inspiring. My grandparents', Kathy, Bob, and Janet, support never faltered and have always motivated me to be the best I can be. My extended family have also been very supportive.

Lastly, I must thank the friends that have stuck with me all these years. April has supported me throughout this journey, despite the distance, and she always believed in me. She kept me down to earth and never failed to tell me the truth, no matter how hard it was, and always made time to listen to any struggles I had. Catie has also supported me throughout the years, despite her own struggles, and I will always admire her perseverance. She is an amazing friend and gave me an outlet for all the stress that comes with grad school. My gratitude for these two ladies cannot be justly expressed in words.

And then there are the furry friends that offer the best laughs and cuddles. Cookie, my childhood cat, always there to comfort me. And Luna and Snickers, who have made living in Boston so far away from home, more bearable.

List of Abbreviations:

abs: absorbance

ADME-tox: absorption, distribution, metabolism, excretion, toxicity

aq: aqueous

ATMA: adamantyltrimethylammonium

b.p.: boiling point

calcd: calculated

CD: circular dichroism

CDK2: cyclin-dependent kinase 2

COSY: correlated spectroscopy

Cy: cyclohexyl

d: doublet

DAD: Diode array detector

Da: Dalton(s)

DART: direct analysis in real time

DCM: dichloromethane

dd: doublet of doublets

d.e.: diastereomeric excess

DMF: *N,N*-dimethylformamide

DMSO: dimethylsulfoxide

d.r.: diastereomeric ratio

dt: doublet of triplets

EAS: electrophilic aromatic substitution

EC_{50}: half maximal effective concentration

e.e.: enantiomeric excess

emi: emission

ESI: electrospray ionization

Et: ethyl

EtOH: ethanol

equiv.: equivalent(s)

eV: electron volts

FASSIF: fasted state simulating intestinal fluid

FT: Fourier Transform

h: hour(s)

HMBC: heteronuclear multiple bond correlation

HOMO: highest occupied molecular orbital

HPLC: high performance liquid chromatography

HRMS: high resolution mass spectrometry

HSQC: heteronuclear single quantum coherence

IC_{50}: half maximal inhibitory concentration

IPTG: Isopropyl-β-D-thiogalactopyranoside

IR: infrared

LC/MS: liquid chromatography/mass spectrometry

LUMO: lowest unoccupied molecular orbital

m: multiplet

M: molar

Me: methyl

m.p.: melting point

NICS: nucleus independent chemical shift

NMQ: *N*-methylquinidine

NMR: nuclear magnetic resonance

NOESY: nuclear overhauser effect spectroscopy

OD_{600}: optical density at 600 nm

PDB: Protein Data Bank

PDE10A: Phosphodiesterase 10A

Pd/C: palladium on carbon

Ph: phenyl

ppm: parts per million

q: quartet

RT: room temperature

s: singlet

STDEV: standard deviation

t: triplet

td: triplet of doublets

TBAF: tetrabutylammonium fluoride

TBE: *tert*-butyl ethylene

TBS: *tert*-butyldimethylsilyl

TBSCl: *tert*-butyldimethylsilyl chloride

THF: tetrahydrofuran

UV: ultraviolet

UV-PES: ultraviolet photoelectron spectroscopy

vis: visible

TABLE OF CONTENTS

Chapter 1: Synthesis and Characterization of a Boron and Nitrogen-containing Tryptophan and its Incorporation into Proteins

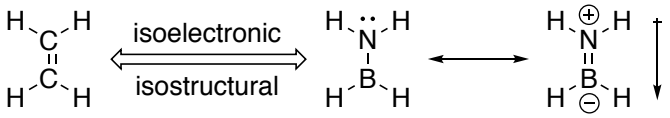
Chapter 1: Synthesis and Characterization of a Boron and Nitrogen-containing Tryptophan and its Incorporation into Proteins

1.1. Introduction

1.1.1. BN/CC Isosterism and Azaborines

The concept of BN/CC isosterism is the isoelectronic and isostructural relationship between a carbon-carbon (CC) bond unit and a boron-nitrogen (BN) bond unit. The effect this isosterism has on a molecule can be shown in one of the simplest cases, ethylene **1.1** and its BN-analogue, aminoborane **1.2** (Figure 1.1).[1] While the number of electrons and general geometry remains the same in each molecule, there are vast differences in their physical, chemical, and electronic properties. While ethylene has no dipole moment, aminoborane has a dipole moment of 1.844 Debye. Ethylene displays a greater bond dissociation energy (BDE) than aminoborane (174.1 kcal/mol versus 139.7 kcal/mol, respectively): the σ-contribution is about the same for both, however ethylene has a greater π-contribution than aminoborane (65 kcal/mol versus 29.9 kcal/mol, respectively). They also display different reactivities. Ethylene is a stable gas under standard conditions with a

Figure 1.1 BN-Isostere of Ethylene



	ethylene **1.1**	aminoborane **1.2**
Dipole	0 D	1.844 D
BDE	174.1 kcal/mol (σ = 109.1, π = 65)	139.7 kcal/mol (σ = 109.8, π = 29.9)

[1] Campbell, P. G.; Marwitz, A. J. V.; Liu, S.-Y. Recent Advances in Azaborine Chemistry. *Angew. Chem. Int. Ed.* **2012**, *51*, 6074–6092.

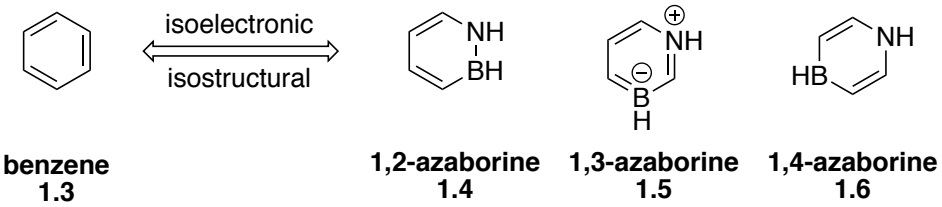
boiling point of –104 °C, while aminoborane is known to be unstable and polymerizes and oligomerizes under standard conditions.

This concept of BN/CC isosterism can be utilized to expand the chemical space of arene-containing compounds that are prominent in medicinal chemistry by replacing a CC bond with a BN bond in benzene, indole, or other arenes.[2] For example, in benzene, there are three possible BN-analogues that contain one boron and one nitrogen: the 1,2-, 1,3-, and 1,4-azaborines (Figure 1.2).[1,3] The inclusion of the BN bond results in a modification of certain electronic characteristics,[4] such as dipole moment,[5] reactivity,[6] and hydrogen

[2] For recent reviews, see: a) McConnell, C. R.; Liu, S.-Y. Late-Stage Functionalization of BN-Heterocycles. *Chem. Soc. Rev.* **2019**, *48*, 3436–3453. b) Giustra, Z. X.; Liu, S.-Y. The State of the Art in Azaborine Chemistry: New Synthetic Methods and Applications. *J. Am. Chem. Soc.* **2018**, *140*, 1184–1194. c) Bélanger-Chabot, G.; Braunschweig, H.; Roy, D. K. Recent Developments in Azaborinine Chemistry. *Eur. J. Inorg. Chem.* **2017**, *2017*, 4353–4368. d) Morgan, M. M.; Piers, W. E. Efficient Synthetic Methods for the Installation of Boron-Nitrogen Bonds in Conjugated Organic Molecules. *Dalt. Trans.* **2016**, *45*, 5920–5924.
[3] Bosdet, M. J.; Piers, W. E. B-N as a C-C Substitute in Aromatic Systems. *Can. J. Chem.* **2009**, *87*, 8–29.
[4] a) McConnell, C. R.; Campbell, P. G.; Fristoe, C. R.; Memmel, P.; Zakharov, L. N.; Li, B.; Darrigan, C.; Chrostowska, A.; Liu, S.-Y. Synthesis and Characterization of 1,2-Azaborine-Containing Phosphine Ligands: A Comparative Electronic Structure Analysis. *Eur. J. Inorg. Chem.* **2017**, *2017*, 2207–2210. b) Liu, Z.; Ishibashi, J. S. A.; Darrigan, C.; Dargelos, A.; Chrostowska, A.; Li, B.; Vasiliu, M.; Dixon, D. A.; Liu, S.-Y. The Least Stable Isomer of BN Naphthalene: Toward Predictive Trends for the Optoelectronic Properties of BN Acenes. *J. Am. Chem. Soc.* **2017**, *139*, 6082–6085. c) Xu, S.; Mikulas, T. C.; Zakharov, L. N.; Dixon, D. A.; Liu, S.-Y. Boron-Substituted 1,3-Dihydro-1,3-Azaborines: Synthesis, Structure, and Evaluation of Aromaticity. *Angew. Chem. Int. Ed.* **2013**, *52*, 7527–7531. d) Campbell, P. G.; Abbey, E. R.; Neiner, D.; Grant, D. J.; Dixon, D. A.; Liu, S.-Y. Resonance Stabilization Energy of 1,2-Azaborines: A Quantitative Experimental Study by Reaction Calorimetry. *J. Am. Chem. Soc.* **2010**, *132*, 18048–18050.
[5] Chrostowska, A.; Xu, S.; Lamm, A. N.; Mazière, A.; Weber, C. D.; Dargelos, A.; Baylère, P.; Graciaa, A.; Liu, S.-Y. UV-Photoelectron Spectroscopy of 1,2- and 1,3-Azaborines: A Combined Experimental and Computational Electronic Structure Analysis. *J. Am. Chem. Soc.* **2012**, *134*, 10279–10285.
[6] a) Lamm, A. N.; Garner, E. B.; Dixon, D. A.; Liu, S.-Y. Nucleophilic Aromatic Substitution Reactions of 1,2-Dihydro-1,2-Azaborine. *Angew. Chem. Int. Ed.* **2011**, *50* , 8157–8160. b) Brown, A. N.; Li, B.; Liu, S.-Y. Negishi Cross-Coupling Is Compatible with a Reactive B-Cl Bond: Development of a Versatile Late-Stage Functionalization of 1,2-Azaborines and Its Application to the Synthesis of New BN Isosteres of Naphthalene and Indenyl. *J. Am. Chem. Soc.* **2015**, *137*, 8932–8935. c) Baggett, A. W.; Vasiliu, M.; Li, B.; Dixon, D. A.; Liu, S.-Y. Late-Stage Functionalization of 1,2-Dihydro-1,2-Azaborines via Regioselective Iridium-Catalyzed C-H Borylation: The Development of a New N,N-Bidentate Ligand Scaffold. *J. Am. Chem. Soc.* **2015**, *137*, 5536–5541. d) Burford, R. J.; Li, B.; Vasiliu, M.; Dixon, D. A.; Liu, S.-Y. Diels-Alder Reactions of 1,2-Azaborines. *Angew. Chem. Int. Ed.* **2015**, *54*, 7823–7827. e) Edel, K.; Yang, X.; Ishibashi, J. S. A.; Lamm, A. N.; Maichle-Mössmer, C.; Giustra, Z. X.; Liu, S.-Y.; Bettinger, H. F. The Dewar Isomer of 1,2-Dihydro-1,2-Azaborinines: Isolation, Fragmentation, and Energy Storage. *Angew. Chem. Int. Ed.* **2018**, *57*, 5296–5300. f) Pan, J.; Kampf, J. W.; Ashe, A. J. Electrophilic Aromatic Substitution Reactions of 1,2-Dihydro-1,2-Azaborines. *Org. Lett.* **2007**, *9*, 679–681.

bonding capabilities,[7] while maintaining the same number of atoms and π-electrons.

Figure 1.2 BN-Isosteres of Benzene

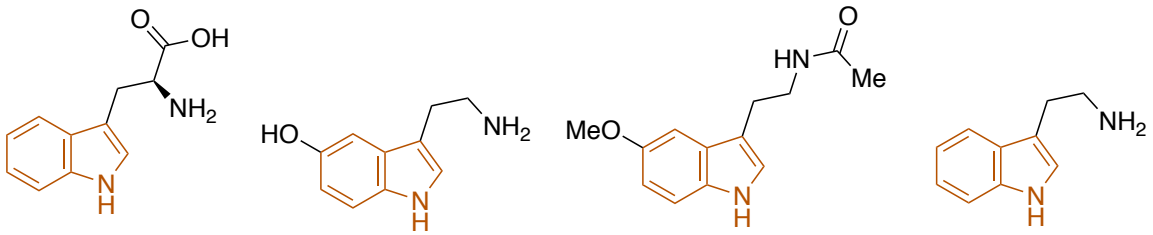


benzene 1.3 | **1,2-azaborine 1.4** | **1,3-azaborine 1.5** | **1,4-azaborine 1.6**

1.1.2. BN/CC Isosterism and Indoles

In addition to monocyclic arenes, BN/CC isosterism can be applied to indole. Indole is a bicyclic nitrogen-containing aromatic heterocycle that is ubiquitous in biological systems and a privileged structure in medicinal chemistry, making it an attractive target for isosterism studies.[8] For example, it is the core of the essential amino acid, tryptophan **1.7**, and the neurotransmitters, serotonin **1.8**, melatonin **1.9**, and tryptamine **1.10** (Figure 1.3).[9] These neurotransmitters are responsible for regulating animal and human sleep cycles, eating habits, and mood.

Figure 1.3 Indole-containing Biologically Active Compounds



Tryptophan 1.7 | **Serotonin 1.8** | **Melatonin 1.9** | **Tryptamine 1.10**

[7] Lee, H.; Fischer, M.; Shoichet, B. K.; Liu, S.-Y. Hydrogen Bonding of 1,2-Azaborines in the Binding Cavity of T4 Lysozyme Mutants: Structures and Thermodynamics. *J. Am. Chem. Soc.* **2016**, *138*, 12021–12024.
[8] For a review, see: Kaushik, N. K.; Kaushik, N.; Attri, P.; Kumar, N.; Kim, C. H.; Verma, A. K.; Choi, E. H. Biomedical Importance of Indoles. *Molecules* **2013**, *18*, 6620–6662.
[9] Fernstrom, J. D. Role of Precursor Availability in Control of Monoamine Biosynthesis in Brain. *Physiol. Rev.* **1983**, *63*, 484–546. b) Gingrich, J. A.; Hen, R. Dissecting the Role of the Serotonin System in Neuropsychiatric Disorders Using Knockout Mice. *Psychopharmacology (Berl).* **2001**, *155*, 1–10. c) Arendt, J. Melatonin: Characteristics, Concerns, and Prospects. *J. Biol. Rhythms* **2005**, *20*, 291–303.

Indole is also a major pharmacophore found in commercially available medications.[10] Just a few examples are shown in Figure 1.4. Ondansetron **1.11**, marketed as Zofran, was developed by GlaxoSmithKline and approved by the FDA in 1991. It is an antiemetic used to treat nausea and vomiting induced by chemotherapy, radiation therapy, and surgery.[11] Sumatriptan **1.12**, marketed as Imitrex, was approved by the FDA in 1992 for the treatment of migraines and was also developed by GlaxoSmithKline.[12] Fluvastatin **1.13**, marketed as Lescol, is used to treat hypercholesterolemia and to prevent cardiovascular disease. Lescol was developed by Novartis and approved for medical use in 1994.[13] Indomethacin **1.14** is a non-steroidal anti-inflammatory drug (NSAID) that was approved in 1965 by the FDA for treatment of fever, pain, and swelling.[14] Osimertinib **1.15**, marketed as Tagrisso, was developed by AstraZeneca and received FDA approval in 2017 for treatment of patients with metastatic non-small cell lung cancer that is positive for the epidermal growth factor receptor (EGFR) T790M mutations.[15] This drug is also listed in the top 200 pharmaceutical drugs by retail sales in 2018.[16] These examples are

[10] For recent reviews, see: a) Naim, M. J.; Alam, O.; Alam, J.; Bano, F.; Alam, P.; Shrivastava, N. Recent Review on Indole: A Privileged Scaffold Structure. *Int. J. Pharma Sci. Res.* **2016**, *7*, 51–62. b) Sravanthi, T. V.; Manju, S. L. Indoles - A Promising Scaffold for Drug Development. *Eur. J. Pharm. Sci.* **2016**, *91*, 1–10. c) Chadha, N.; Silakari, O. Indoles as Therapeutics of Interest in Medicinal Chemistry: Bird's Eye View. *Eur. J. Med. Chem.* **2017**, *134*, 159–184.

[11] Scuderi, P.; Wetchler, B.; Sung, Y.-F.; Mingus, M.; DuPen, S.; Claybon, L.; Leslie, J.; Talke, P.; Apfelbaum, J.; Sharifi-Azad, S. Treatment of Postoperative Nausea and Vomiting after Outpatient Surgery with the 5-HT3 Antagonist Ondansetron. *Anesthesiology* **1993**, *78*, 15–20.

[12] Ferrari, et al. Treatment of Migraine Attacks with Sumatriptan. *N. Engl. J. Med.* **1991**, *325*, 316–321.

[13] Jones, P.; Kafonek, S.; Laurora, I.; Hunninghake, D. Comparative Dose Efficacy Study of Atorvastatin Versus Simvastatin, Pravastatin, Lovastatin, and Fluvastatin in Patients With Hypercholesterolemia (The CURVES Study). *Am. J. Cardiol.* **1998**, *81*, 582–587.

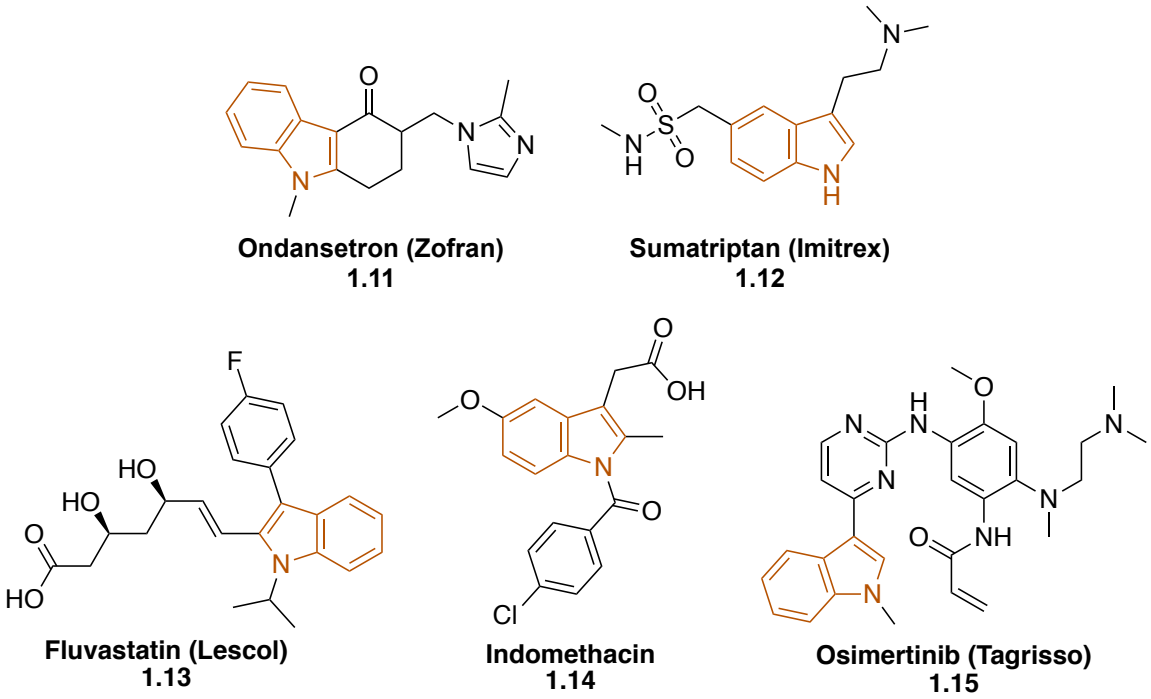
[14] Hart, F. D.; Boardman, P. L. Indomethacin a New Non-Steroid Anti-Inflammatory Agent. *Br. Med. J.* **1963**, *2*, 965–970.

[15] Greig, S. L. Osimertinib: First Global Approval. *Drugs* **2016**, *76*, 263–273.

[16] McGrath, N. A.; Brichacek, M.; Njardarson, J. T. A Graphical Journey of Innovative Organic Architectures That Have Improved Our Lives. *J. Chem. Educ.* **2010**, *87*, 1348–1349.

only a few of the many different indole containing compounds that are used to treat various conditions.

Figure 1.4 Selected Examples of Indole-containing Pharmaceuticals on the Market

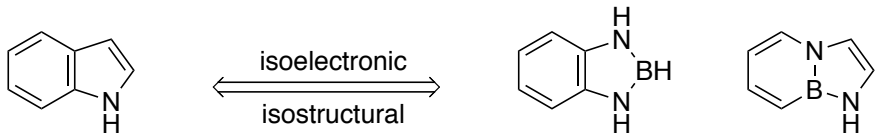


Two BN-isosteres of indole have been developed thus far: the "external" BN-indole **1.17** and "fused" BN-indole **1.18** (Figure 1.5). Goubeau and coworkers first introduced the synthesis of a substituted "external" BN-indole in 1957[17] and later reported the synthesis of the parent compound **1.17** in 1964.[18] Since then, other studies have emerged on its synthesis, applications, and properties.[19] It has mostly found application as a boryl-ligand in organometallic chemistry and aromatic building block in optoelectronic materials.[20]

---

[17] Ulmschneider, D.; Goubeau, J. Reaktionen des Trimethylbors. *Chem. Ber.* **1957**, *90*, 2733–2738.
[18] Goubeau, J.; Snyder, H. Darstellung von Diaminoborinen aus Diamin-monohydrochloriden und Natriumborhydrid. *Justus Liebigs Ann. Chem.* **1964**, *675*, 1–9.
[19] For recent reviews, see: a) Abbey, E. R.; Liu, S.-Y. BN Isosteres of Indole. *Org. Biomol. Chem.* **2013**, *11*, 2060–2069. b) Weber, L.; Böhling, L. The Role of 2,3-Dihydro-1-H-1,3,2-Diazaboroles in Luminescent Molecules. *Coord. Chem. Rev.* **2015**, *284*, 236–275.
[20] For recent examples, see: a) Weber, L.; Halama, J.; Böhling, L.; Brockhinke, A.; Chrostowska, A.; Darrigan, C.; Dargelos, A.; Stammler, H. G.; Neumann, B. 1,3,2-Benzodiazaboroles With 1,3-Pentafluorophenyl and Tetrafluoropyridyl Substituents As Building Blocks in Luminescent Compounds. *Eur. J. Inorg. Chem.* **2013**, *24,* 4268–4279. b) Weber, L.; Halama, J.; Hanke, K.; Böhling, L.; Brockhinke,

Figure 1.5 BN-Isosteres of Indole



Indole 1.16 | **"External" BN-Indole 1.17** | **"Fused" BN-Indole 1.18**

More recently, the N-substituted "fused" BN-indole was introduced by the Liu lab in 2010,[21] with the parent variant **1.18** appearing a year later in 2011.[22] When **1.18** is compared with the natural indole **1.16**, it is found to have an almost identical structure and the same electrophilic aromatic substitution reactivity, albeit with increased nucleophilicity at the C3 position.[23] Additionally, it was found that the N-H of "fused" BN-indole **1.18** has a p*K*a of 30 (versus 21 for indole **1.16**), making it less acidic than natural indole **1.16**.

While the "external" BN-indole **1.17** has a multitude of applications as boryl-ligands and as a component in optoelectronic materials, its hydrolytic instability has prevented its use in biomedical research. On the other hand, the "fused" BN-indole **1.18** has been determined to be relatively stable in water. The physical and chemical features of the "fused" BN-indole could potentially be utilized in biomedical studies to alter the

A.; Stammler, H. G.; Neumann, B.; Fox, M. A. On the Ambiguity of 1,3,2-Benzodiazaboroles as Donor/Acceptor Functionalities in Luminescent Molecules. *Dalt. Trans.* **2014**, *43*, 3347–3363. c) Davies, G. H. M.; Molander, G. A. Synthesis of Functionalized 1,3,2-Benzodiazaborole Cores Using Bench-Stable Components. *J. Org. Chem.* **2016**, *81*, 3771–3779. d) Kwan, E. H.; Kawai, Y. J.; Kamakura, S.; Yamashita, M. A Long-Tethered (P-B-P)-Pincer Ligand: Synthesis, Complexation, and Application to Catalytic Dehydrogenation of Alkanes. *Dalt. Trans.* **2016**, *45*, 15931–15941. e) Fu, Y.; Zhang, K.; Dmitrieva, E.; Liu, F.; Ma, J.; Weigand, J. J.; Popov, A. A.; Berger, R.; Pisula, W.; Liu, J.; et al. NBN-Embedded Polycyclic Aromatic Hydrocarbons Containing Pentagonal and Heptagonal Rings. *Org. Lett.* **2019**, *21*, 1354–1358.

[21] Abbey, E. R.; Zakharov, L. N.; Liu, S.-Y. Electrophilic Aromatic Substitution of a BN Indole. *J. Am. Chem. Soc.* **2010**, *132*, 16340–16342.

[22] Abbey, E. R.; Zakharov, L. N.; Liu, S.-Y. Boron in Disguise: The Parent Fused BN Indole. *J. Am. Chem. Soc.* **2011**, *133*, 11508–11511.

[23] Chrostowska, A.; Xu, S.; Mazière, A.; Boknevitz, K.; Li, B.; Abbey, E. R.; Dargelos, A.; Graciaa, A.; Liu, S.-Y. UV-Photoelectron Spectroscopy of BN Indoles: Experimental and Computational Electronic Structure Analysis. *J. Am. Chem. Soc.* **2014**, *136*, 11813–11820.

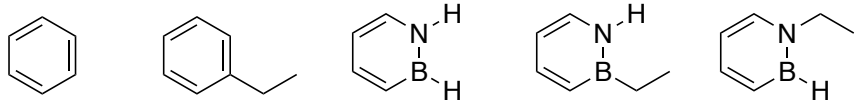
properties of known, biologically active, indole-containing compounds with minimal disturbance to the size of the molecule.

1.2. Background

1.2.1. 1,2-azaborines in Biological Systems

While the chemistry of 1,2-azaborines has been expanded upon dramatically in recent years, little is known about the interactions of this class of arenes in the context of biological systems. The Liu group has led efforts into uncovering the potential for azaborine application to biological systems and has investigated the binding of 1,2-azaborines in polar and non-polar binding sites in biological macromolecules.[2b]

Figure 1.6 Ligand Binding Study Substrates

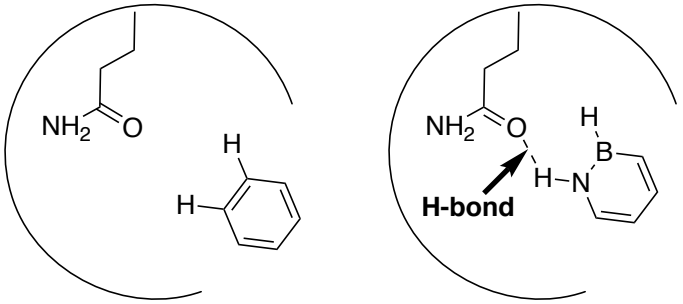


In the first such report in 2009, Liu and coworkers demonstrated through protein crystal structure analysis that 1,2-azaborines **1.4**, **1.20**, and **1.21** bind to the engineered nonpolar cavity in the L99A T4 lysozyme mutant in a similar orientation as the carbonaceous arene substrates **1.3** and **1.19** (Figure 1.6).[24] More recently, in 2016, the binding of 1,2-azaborines was explored in an engineered polar cavity in the L99A/M102Q T4 lysozyme double mutant in order to study the hydrogen bonding capabilities of the azaborine N–H bond.[7] Liu showed through protein crystal structure analysis of the bound azaborines that a hydrogen bond is present between the N-H of the azaborine and the

[24] Liu, L.; Marwitz, A. J. V.; Matthews, B. W.; Liu, S.-Y. Boron Mimetics: 1,2-Dihydro-1,2-Azaborines Bind inside a Nonpolar Cavity of T4 Lysozyme. *Angew. Chem. Int. Ed.* **2009**, *48*, 6817–6819.

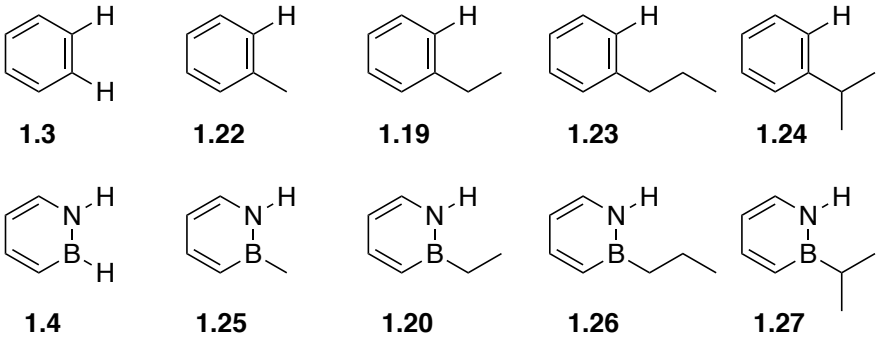
carbonyl of the glutamine in the binding pocket, an interaction which is absent in the carbonaceous compound (Figure 1.7).

Figure 1.7 Crystal Structure Representation of Benzene and Azaborine Binding in the Polar Binding Pocket of T4 Lysozyme



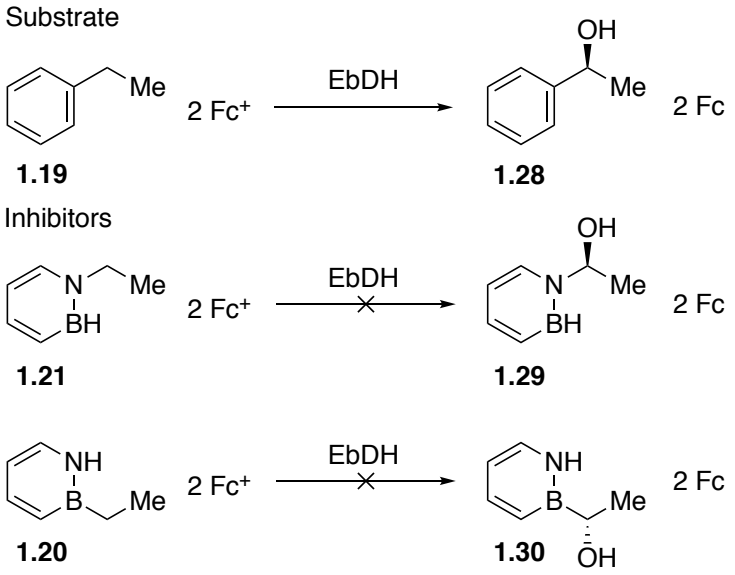
The strengths of these hydrogen bonds were also measured through the use of isothermal titration calorimetry (ITC) and a double mutant cycle analysis. The parent azaborine hydrogen bond was measured to be ~1 kcal/mol while the *B*-ethyl-azaborine hydrogen bond was ~0.6 kcal/mol. In 2019, another report from the Liu group demonstrated the effect sterics had on the hydrogen bond strength (Figure 1.8).[25] This was done through a similar study where the substituent on the boron was varied and the hydrogen bond strength measured. It was found that increasing the steric profile of the boron substituent decreased the strength of the hydrogen bond in a discreet fashion.

Figure 1.8 Ligand Binding Substrates in Hydrogen Bond Strength Study



[25] Liu, Y.; Liu, S.-Y. Exploring the Strength of a Hydrogen Bond as a Function of Steric Environment Using 1,2-Azaborine Ligands and Engineered T4 Lysozyme Receptors. *Org. Biomol. Chem.* **2019**, *17*, 7002–7006.

Figure 1.9 Oxidation of Ethylbenzene to (*S*)-1-Phenylethanol



In 2013, the Liu and Heider labs reported that both *N*-ethyl and *B*-ethyl-1,2-azaborines inhibit the activity of ethylbenzene dehydrogenase (EbDH), an enzyme which catalyzes the hydroxylation of ethylbenzene to (*S*)-1-phenylethanol (Figure 1.9).[26] It was calculated to show that a possible reason for this variation in activity between the sterically similar compounds is the result of the differences in stabilities of the benzylic radical and benzylic cation, both of which are believed to be intermediates in the oxidation of ethylbenzene (Table 1.1). The relative instability of the radical and cationic intermediates

Table 1.1 Relative Calculated Energy Values for Radical and Carbocation Formation

Compound	ΔΔG radical (kJ/mol)	ΔΔG carbocation (kJ/mol)
ethylbenzene **1.19**	0	0
N-ethyl-1,2-azaborine **1.21**	23.69	12.28
B-ethyl-1,2-azaborine **1.20**	15.41	24.43

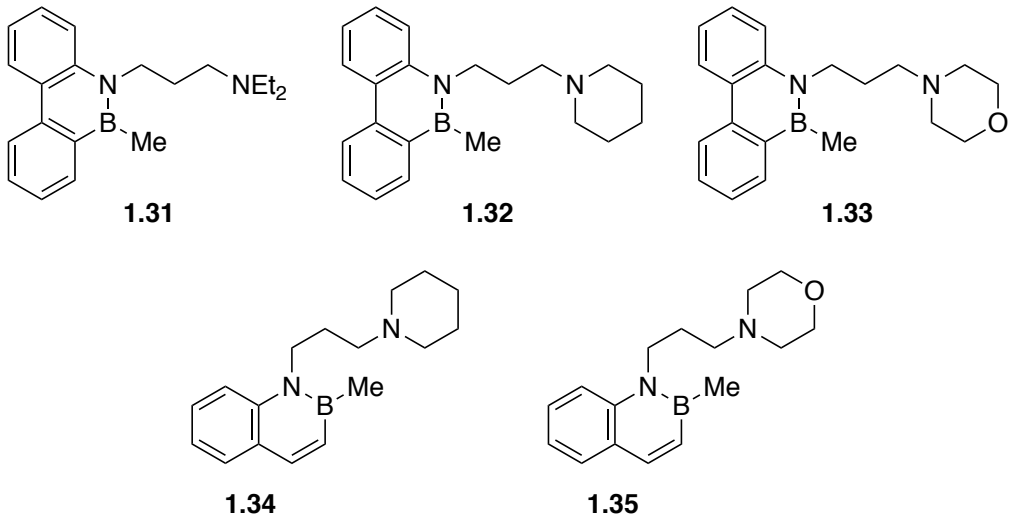
[26] Knack, D. H.; Marshall, J. L.; Harlow, G. P.; Dudzik, A.; Szaleniec, M. I.; Liu, S.-Y.; Heider, J. BN/CC Isosteric Compounds as Enzyme Inhibitors: N- and B-Ethyl-1,2-Azaborine Inhibit Ethylbenzene Hydroxylation as Nonconvertible Substrate Analogues. *Angew. Chem. Int. Ed.* **2013**, *52*, 2599–2601.

associated with 1,2-azaborine is believed to be the cause for their observed inhibitory activity against EbDH. This observed role reversal between carbonaceous arenes and 1,2-azaborines is a direct result of BN/CC isosterism.

1.2.2. 1,2-azaborines in Medicinal Chemistry

Arenes are a ubiquitous pharmacophore found in many pharmaceuticals.[16] The application of BN/CC isosterism in the field of medicinal chemistry expands the structural possibilities for new medications. One of the first examples of an application of BN-containing heterocycles to medicinal chemistry was in Dewar's work in 1964.[27] Dewar synthesized multiple water-soluble BN-phenanthrene and BN-naphthalene isosteres and tested their viability as reagents in boron neutron capture therapy (BNCT) (Figure 1.10). While they were found to be very toxic when introduced to mice, the results indicated the compounds were stable *in vivo*.

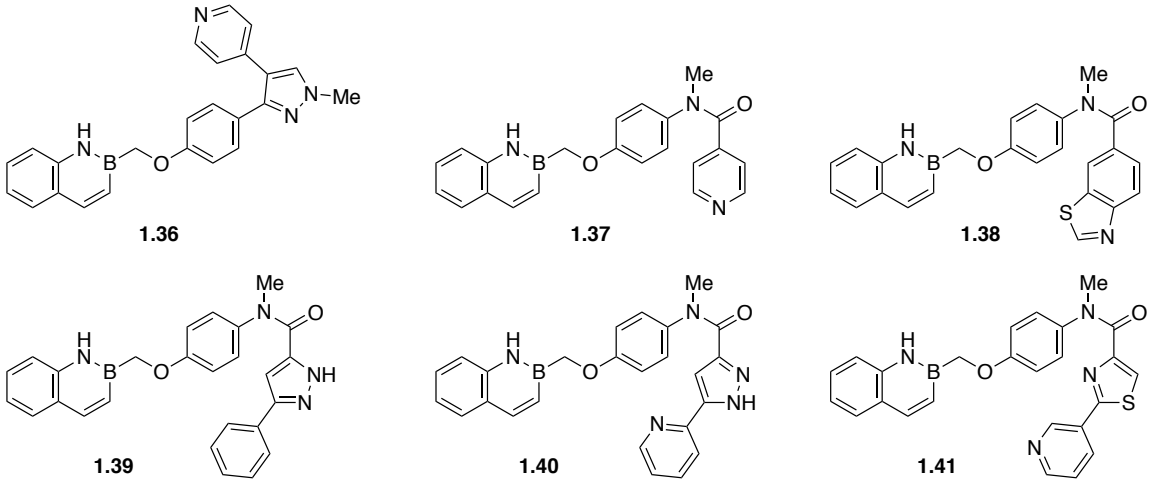
Figure 1.10 BN-Isosteres of Naphthalene and Phenanthrene-containing Drug-like Molecules Synthesized for BNCT Studies.



[27] Dewar, M. J. S.; Hashmall, J.; Kubba, V. P. New Heteroaromatic Boron Compounds. XIX. Water-Soluble Derivatives of 10,9-Borazarophenanthrene and 2,1-Borazaronaphthalene as Potential Agents for Neutron Capture Therapy. *J. Org. Chem.* **1964**, *29*, 1755–1757.

It wasn't until 2015 that another application of BN-containing arene isosteres in drug-like molecules was reported. Kilburn and coworkers synthesized and studied multiple BN-containing analogues of MP10, the compound used as a benchmark comparison of PDE10A inhibition, as well as other PDE10A inhibitors, with the intent to gain a better understanding of how the similarity between physical properties of the isosteres would correlate to similarities in biological interactions (Figure 1.11).[28] They found that the activity of borazaronaphthyl-substituted analogues was diminished from the benchmark compound, but were nonetheless more potent than the naphthyl-substituted analogues. It was also shown that there is some degree of interaction with the active site, and, therefore, there is potential for the use of BN-naphthalenes in future medicinal chemistry studies.

Figure 1.11 BN-Isosteres of PDE10A inhibitors

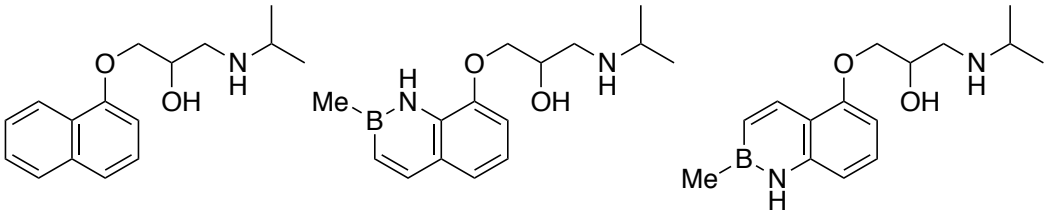


Later that year, Rombouts and coworkers synthesized two BN-naphthalene-containing analogues of propranolol, a naphthalene-containing β-blocker drug, and compared their properties *in vitro* and *in vivo* (Figure 1.12).[29] Here they found that the

[28] Vlasceanu, A.; Jessing, M.; Kilburn, J. P. BN/CC Isosterism in Borazaronaphthalenes towards Phosphodiesterase 10A (PDE10A) Inhibitors. *Bioorg. Med. Chem.* **2015**, *23*, 4453–4461.
[29] Rombouts, F. J. R.; Tovar, F.; Austin, N.; Tresadern, G.; Trabanco, A. A. Benzazaborinines as Novel Bioisosteric Replacements of Naphthalene: Propranolol as an Example. *J. Med. Chem.* **2015**, *58*, 9287–9295.

azaborine analogue **1.44** showed similar inhibitory potency against the $\beta 2$ receptor and similar selectivity profile to propranolol **1.42**, as well as a comparable *in vitro* ADME-tox profile.

Figure 1.12 Propranolol and two BN-isosteres



1.42 **1.43** **1.44**

Lastly, in 2017, three BN-isosteres of known biologically active compounds were synthesized and tested *in vivo* (Figure 1.13).[30] It was found that in all cases, the BN-analogues showed similar activity to their corresponding CC-analogues as well as increased aqueous solubility *in vitro*. Of particular note, the BN-analogue **1.47b** of the CDK2 inhibitor **1.47a** showed an IC_{50} of 87 nM, which is significantly more potent than the IC_{50} of 320 nM for the CC-analogue. This could be attributed to the additional hydrogen-bonding interaction of the azaborine N–H with the binding site of CDK2. Further testing in a rat model demonstrated the BN-analogue has superior pharmacokinetic properties, as shown in Table 1.2. The greater bioavailability and lower clearance result in a two-fold increase in the AUC_{po} (area under the curve per oral administration). In conclusion, these desirable properties are the result of the greater solubility of the azaborine compound in physiological conditions.

[30] Zhao, P.; Nettleton, D. O.; Karki, R. G.; Zécri, F. J.; Liu, S.-Y. Medicinal Chemistry Profiling of Monocyclic 1,2-Azaborines. *ChemMedChem* **2017**, *12*, 358–361.

Figure 1.13 The Effect of BN/CC Isosterism on the Biological Activity of Pharmaceutically Active Compounds

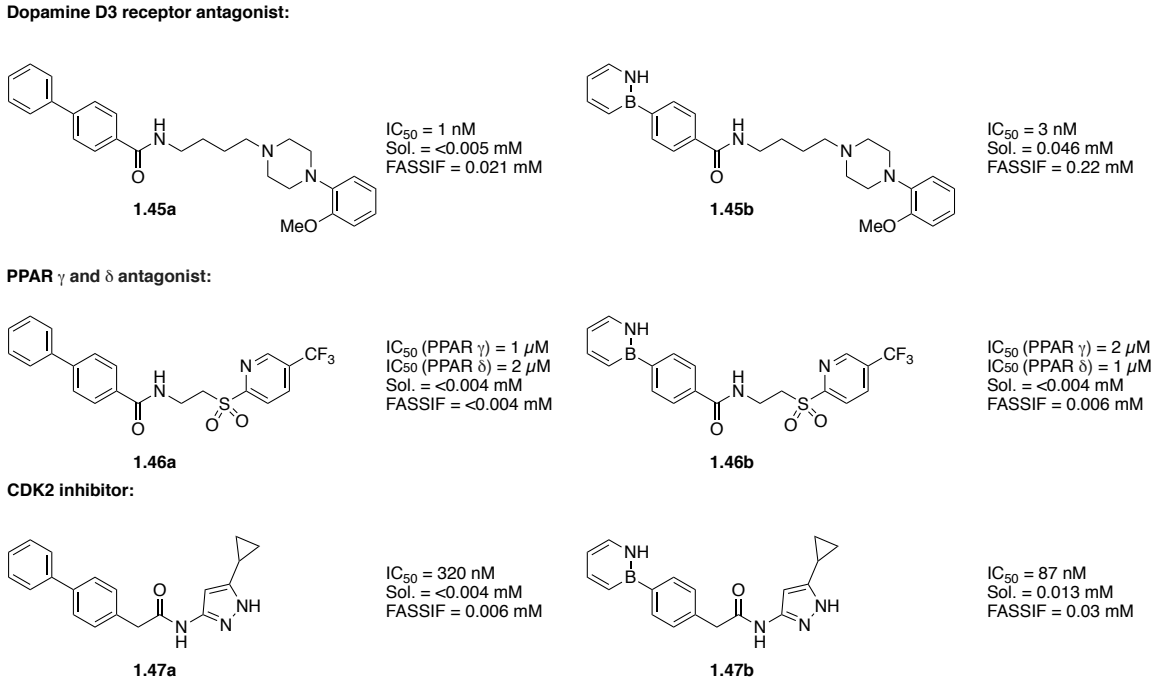


Table 1.2 Pharmacokinetic Properties of CDK2 Inhibitors

Compd	i.v. [mg/kg][a]	p.o. [mg/kg][b]	AUC_{iv} [nM·h][c]	AUC_{po} [nM·h][d]	*F* [%][e]	CL [mL/min·kg][f]	$t_{1/2}$ [h]	t_{max} [h]	MRT [h][g]	C_{max} [nM]
1.47a	0.5	5.0	1261	283	22.5	39.2	10.5	0.5	0.8	692
1.47b	1.0	5.0	2092	613	29.3	23.7	11.6	1.5	2.7	746

[a] Intravenous dose. [b] Per os dose. [c] Area under the curve (intravenous) normalized to 1 mg/kg dose. [d] Area under the curve (oral) normalized to 1 mg/kg dose. [e] Bioavailability. [f] Clearance rate. [g] Mean resonance time.

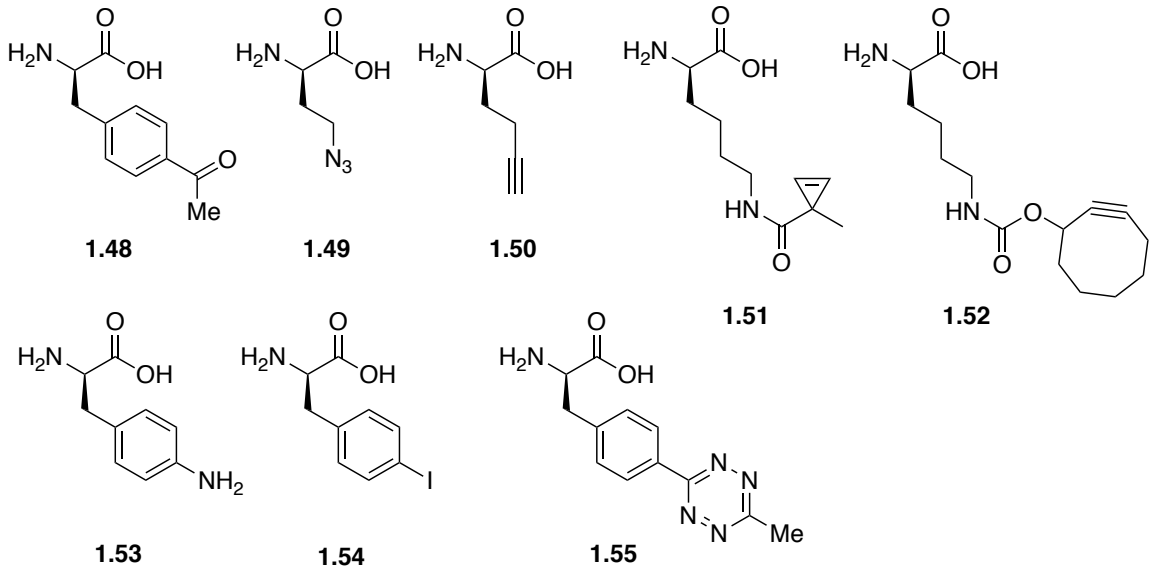
1.2.3. Unnatural Analogues of Proteinogenic Amino Acids

Proteins in all living things are made up of 20 amino acids (with the rare exception of selenocysteine and pyrrolysine).[31] While nature is able to function on these 20 amino acids, biochemists have determined a need to expand the selection of functional groups available in order to study proteins. Unnatural amino acids (UAAs) have been developed

[31] Ambrogelly, A.; Palioura, S.; Söll, D. Natural Expansion of the Genetic Code. *Nat. Chem. Biol.* **2007**, *3*, 29–35.

for this purpose. Being able to incorporate new functionalities into proteins can offer ways to explore and engineer the function, properties, and stability of proteins.[32]

Figure 1.14 Selection of UAAs used in Protein Studies



Researchers have found and developed many applications for UAAs. Some UAAs have been used to incorporate reactive handles into proteins for bioconjugation reactions (Figure 1.14).[33] Numerous other UAAs have been developed as endogenous amino acid analogues to alter the fluorescent character of proteins with minimal perturbation to the native structure of the protein.[34] Others still have been developed to probe the cellular

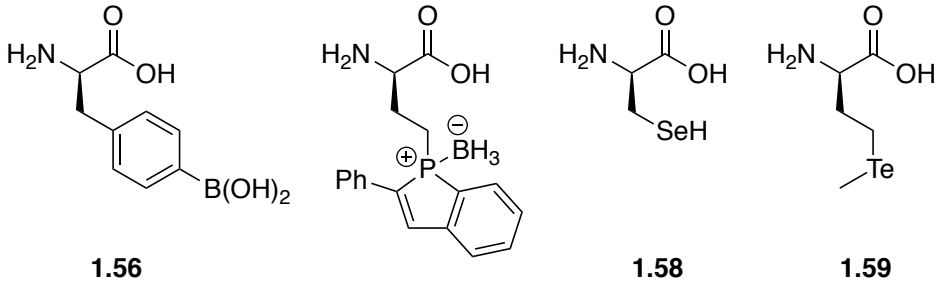
[32] a) Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444. b) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing Logical Codon Reassignment-Expanding the Chemistry in Biology. *Chem. Sci.* **2015**, *6*, 50–69. c) Neumann-Staubitz, P.; Neumann, H. The Use of Unnatural Amino Acids to Study and Engineer Protein Function. *Curr. Opin. Struct. Biol.* **2016**, *38*, 119–128.

[33] a) Kim, C. H.; Axup, J. Y.; Schultz, P. G. Protein Conjugation with Genetically Encoded Unnatural Amino Acids. *Curr. Opin. Chem. Biol.* **2013**, *17*, 412–419. b) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114*, 4764–4806. c) Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*, 1–14.

[34] Ross, J. B. A.; Szabo, A. G.; Hogue, C. W. V. Enhancement of Protein Spectra with Tryptophan Analogs: Fluorescence Spectroscopy of Protein-Protein and Protein-Nuclei Acid Interactions. *Methods Enzymol.* **1997**, *278*, 151–190.

mechanisms of protein synthesis.[35] Some UAAs have been developed to include elements not commonly found in natural amino acids, such as boron[36], phosphorous[37], selenium[38], and tellurium[39,38b] for use as therapeutics (e.g. **1.56**), fluorescent tags (e.g. **1.57**), handles for bioconjugation (e.g. **1.58**), or to aid in protein structure determination (e.g. **1.59**) (Figure 1.15).

Figure 1.15 Selection of UAAs used in Protein Studies



1.56 **1.57** **1.58** **1.59**

[35] Ngo, J. T.; Tirrell, D. A. Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis. *Acc. Chem. Res.* **2011**, *44*, 677–685.

[36] a) Diaz, D. B.; Yudin, A. K. The Versatility of Boron in Biological Target Engagement. *Nat. Chem.* **2017**, *9*, 731–742. b) Das, B. C.; Thapa, P.; Karki, R.; Schinke, C.; Das, S.; Kambhampati, S.; Banerjee, S. K.; Van Veldhuizen, P.; Verma, A.; Weiss, L. M.; Evans, T. Boron Chemicals in Diagnosis and Therapeutics. *Future Med. Chem.* **2013**, *5*, 653–676.

[37] Arribat, M.; Rémond, E.; Clément, S.; Lee, A. Van Der; Cavelier, F. Phospholyl(Borane) Amino Acids and Peptides: Stereoselective Synthesis and Fluorescent Properties with Large Stokes Shift. *J. Am. Chem. Soc.* **2018**, *140*, 1028–1034.

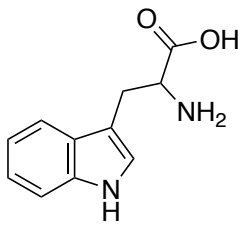
[38] a) Cohen, D. T.; Zhang, C.; Fadzen, C. M.; Mijalis, A. J.; Hie, L.; Johnson, K. D.; Shriver, Z.; Plante, O.; Miller, S. J.; Buchwald, S. L.; Pentelute, B. L. A Chemoselective Strategy for Late-Stage Functionalization of Complex Small Molecules with Polypeptides and Proteins. *Nat. Chem.* **2019**, *11*, 78–85. b) Satheeshkumar, K.; Raju, S.; Singh, H. B.; Butcher, R. J. Reactivity of Selenocystine and Tellurocystine: Structure and Antioxidant Activity of the Derivatives. *Chem. Eur. J.* **2018**, *24*, 17513–17522. c) Mukai, T.; Sevostyanova, A.; Suzuki, T.; Fu, X.; Söll, D. A Facile Method for Producing Selenocysteine-Containing Proteins. *Angew. Chem. Int. Ed.* **2018**, *57*, 7215–7219. d) Dantas De Araujo, A.; Perry, S. R.; Fairlie, D. P. Chemically Diverse Helix-Constrained Peptides Using Selenocysteine Crosslinking. *Org. Lett.* **2018**, *20*, 1453–1456.

[39] Budisa, N.; Karnbrock, W.; Steinbacher, S.; Humm, A.; Prade, L.; Neuefeind, T.; Moroder, L.; Huber, R. Bioincorporation of Telluromethionine into Proteins: A Promising New Approach for X-Ray Structure Analysis of Proteins. *J. Mol. Biol.* **1997**, *270*, 616–623.

1.2.4. Unnatural Tryptophan Analogues and Their Applications

Tryptophan residues serve as the target of many protein studies. Tryptophan **1.60** (Figure 1.16) is the main source of intrinsic fluorescence in proteins,[40] participates in various π-system interactions,[41] and accounts for about 1% of amino acids found in proteins.[42] Tryptophan's relative scarcity, role in protein stability, and utility as a spectroscopic handle mark it as a useful target for protein studies. As such, many tryptophan analogues have been developed and studied (Figure 1.17).

Figure 1.16 Trytophan



1.60

Two of the first unnatural tryptophan analogues, 7-azatryptophan **1.60a**[43] and tryptazan **1.60b**[44], were incorporated into proteins in *E. coli* in the 1950's. When later incorporated into the enzyme alkaline phosphatase, it was found that the enzyme did not lose its activity and the only evidence of incorporation was a change in the absorbance and emission spectra.[45] Since then, many other tryptophan analogues with different functional groups have been incorporated for various purposes (Figure 1.17). Other azatryptophans (**1.60c-d**), as well as many other functional group-containing tryptophans (**1.60e-h**), have

---

[40] Callis, P. R. 1L(a) and 1L(b) Transitions of Tryptophan: Applications of Theory and Experimental Observations to Fluorescence of Proteins. *Methods Enzymol.* **1997**, *278*, 113–150.

[41] a) Ruan, C.; Rodgers, M. T. Cation-π Interactions: Structures and Energetics of Complexation of Na^+ and K^+ with the Aromatic Amino Acids, Phenylalanine, Tyrosine, and Tryptophan *J. Am. Chem. Soc.* **2004**, *126*, 14600–14610. b) Dougherty, D. a. Cation-π Interactions Involving Aromatic Amino Acids *J. Nutr.* **2007**, *137*, 1504–1508. c) Dougherty, D. A. Physical Organic Chemistry on the Brain. *J. Org. Chem.* **2008**, *73*, 3667–3673. d) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* 1**985**, *229,* 23–28.

[42] McCaul, C. P.; Ludescher, R. D. Room Temperature Phosphorescence from Tryptophan and Halogenated Tryptophan Analogs in Amorphous Sucrose. *Photochem. Photobiol.* **1999**, *70*, 166–171.

[43] Pardee, A. B.; Shore, V. G.; Prestidge, L. S. Incorporation of Azatryptophan into Proteins of Bacteria and Bacteriophage. *Biochim. Biophys. Acta* **1956**, *21*, 406–407.

[44] Brawerman, G.; Yčas, M. Incorporation of the Amino Acid Analog Tryptazan into the Protein of Escherichia Coli. *Arch. Biochem. Biophys.* **1957**, *68*, 112–117.

[45] Attias, J.; Schlesinger, M. J.; Schlesinger, S. The Effect of Amino Acid Analogues on Alkaline Phosphatase Formation in Escherichia Coli K-12. *J. Biol. Chem.* **1969**, *244*, 3810–3817.

been used to alter the spectroscopic properties.[46] Fluorinated tryptophans have been used to suppress fluorescence (**1.60i**),[47] alter the indole's capability to participate in intermolecular forces and affect protein stability (**1.60i-o**),[48] and for use in ^{19}F NMR studies (**1.60i-k**).[49] Tryptophans functionalized with amino groups have been used as pH sensors (**1.60p-q**)[50] and, along with other analogues (**1.60r-x**), to alter and study the spectroscopic properties of fluorescent proteins.[51] Sulfur and selenium containing analogues (**1.60y-ab**) have been used to both alter fluorescence and as a handle in x-ray crystallography.[52] Others

---

[46] a) Lepthien, S.; Hoesl, M. G.; Merkel, L.; Budisa, N. Azatryptophans Endow Proteins with Intrinsic Blue Fluorescence. *Proc. Natl. Acad. Sci.* **2008**, *105*, 16095–16100. b) Talukder, P.; Chen, S.; Arce, P. M.; Hecht, S. M. Efficient Asymmetric Synthesis of Tryptophan Analogues Having Useful Photophysical Properties. *Org. Lett.* **2014**, *16*, 556–559. c) Talukder, P.; Chen, S.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. Cyanotryptophans as Novel Fluorescent Probes for Studying Protein Conformational Changes and DNA-Protein Interaction. *Biochemistry* **2015**, *54*, 7457–7469.

[47] Bronskill, P. M.; Wong, J. T. Suppression of Fluorescence of Tryptophan Residues in Proteins by Replacement with 4-Fluorotryptophan. *Biochem. J.* **1988**, *249*, 305–308.

[48] a) Minks, C.; Huber, R.; Moroder, L.; Budisa, N. Atomic Mutations at the Single Tryptophan Residue of Human Recombinant Annexin V: Effects on Structure, Stability, and Activity. *Biochemistry* **1999**, *38*, 10649–10659. b) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. From Ab Initio Quantum Mechanics to Molecular Neurobiology: A Cation-π Binding Site in the Nicotinic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12088–12093. c) Tobola, F.; Lelimousin, M.; Varrot, A.; Gillon, E.; Darnhofer, B.; Blixt, O.; Birner-Gruenberger, R.; Imberty, A.; Wiltschi, B. Effect of Noncanonical Amino Acids on Protein-Carbohydrate Interactions: Structure, Dynamics, and Carbohydrate Affinity of a Lectin Engineered with Fluorinated Tryptophan Analogs. *ACS Chem. Biol.* **2018**, *13*, 2211–2219.

[49] Seifert, M. H. J.; Ksiazek, D.; Azim, M. K.; Smialowski, P.; Budisa, N.; Holak, T. A. Slow Exchange in the Chromophore of a Green Fluorescent Protein Variant. *J. Am. Chem. Soc.* **2002**, *124*, 7932–7942.

[50] Sinha, H. K.; Dogra, S. K.; Krishnamurthy, M. Excited-State and Ground-State Proton-Transfer Reactions in 5-Aminoindole. *Bull. Chem. Soc. Jpn*. **1987**, *60*, 4401–4407. b) Budisa, N.; Rubini, M.; Bae, J. H.; Weyher, E.; Wenger, W.; Golbik, R.; Huber, R.; Moroder, L. Global Replacement of Tryptophan with Aminotryptophans Generates Non-Invasive Protein-Based Optical PH Sensors. *Angew. Chem. Int. Ed.* **2002**, *41*, 4066–4069.

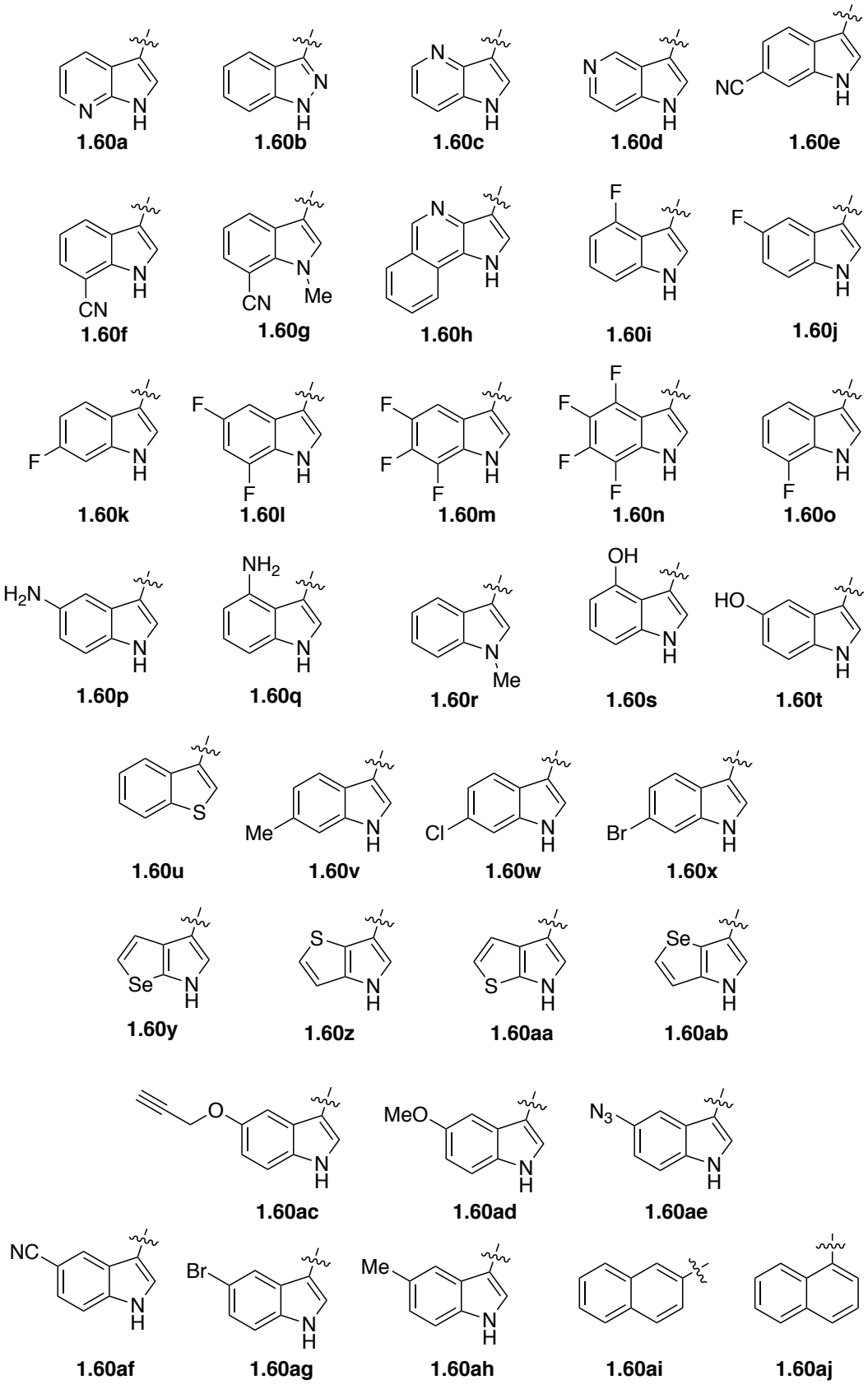
[51] a) Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H. J.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N. Expansion of the Genetic Code Enables Design of a Novel "Gold" Class of Green Fluorescent Proteins. *J. Mol. Biol.* **2003**, *328*, 1071–1081. b) Chatterjee, A.; Xiao, H.; Yang, P. Y.; Soundararajan, G.; Schultz, P. G. A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in E.Coli. *Angew. Chem. Int. Ed.* **2013**, *52*, 5106–5109. c) Kwon, I.; Tirrell, D. A. Site-Specific Incorporation of Tryptophan Analogues into Recombinant Proteins in Bacterial Cells. *J. Am. Chem. Soc.* **2007**, *129*, 10431–10437.

[52] Bae, J. H.; Alefelder, S.; Kaiser, J. T.; Friedrich, R.; Moroder, L.; Huber, R.; Budisa, N. Incorporation of β-Selenolo[3,2-b]Pyrrolyl-Alanine into Proteins for Phase Determination in Protein X-Ray Crystallography. *J. Mol. Biol.* **2001**, *309*, 925–936. b) Budisa, N.; Alefelder, S.; Bae, J. H.; Golbik, R.; Minks, C.; Huber, R.; Moroder, L. Proteins with β-(Thienopyrrolyl)Alanines as Alternative Chromophores and Pharmaceutically Active Amino Acids. *Protein Sci.* **2001**, *10*, 1281–1292. c) Budisa, N.; Pal, P. P.; Alefelder, S.; Birle, P.; Krywcun, T.; Rubini, M.; Wenger, W.; Bae, J. H.; Steiner, T. Probing the Role of Tryptophans in Aequorea Victoria Green Fluorescent Proteins with an Expanded Genetic Code. *Biol. Chem.* **2004**, *385*, 191–202.

(**1.60t,ac-ae**) have been incorporated with a reactive handle for bioconjugation reactions.[53] Lastly, various others have been successfully incorporated into proteins to study ligand binding and spectroscopic properties (**1.60af-aj**).[48b]

[53] a) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **2017**, *139*, 11670–11673. b) Italia, J. S.; Addy, P. S.; Wrobel, C. J. J.; Crawford, L. A.; Lajoie, M. J.; Zheng, Y.; Chatterjee, A. An Orthogonalized Platform for Genetic Code Expansion in Both Bacteria and Eukaryotes. *Nat. Chem. Biol.* **2017**, *13*, 446–450. c) Addy, P. S.; Italia, J. S.; Chatterjee, A. An Oxidative Bioconjugation Strategy Targeted to a Genetically Encoded 5-Hydroxytryptophan. *ChemBioChem* **2018**, *19*, 1375–1378.

Figure 1.17 Unnatural Analogues of Tryptophan



1.2.5. Methods of Incorporation of Unnatural Amino Acid Analogues

Significant efforts have gone into expanding the genetic code and incorporating unnatural amino acids (UAAs) into proteins and peptides.[54] There are multiple ways of incorporating UAAs into proteins, both *in vitro* and *in vivo*, and the utility of different methods depends on the application. In the case of shorter peptides, *in vitro* methods such as solid-phase peptide synthesis, native chemical ligation,[55] and *in vitro* translation[56] are the best options. For larger proteins, these methods can result in low yields and are technically very challenging. They also do not solve the problem of incorporating a UAA into a protein within the cell. There are two main ways for incorporating proteins *in vivo via* protein translation: residue-specific methods and site-specific methods, both of which use the cell's endogenous biosynthetic machinery.

The residue-specific method utilizes the aminoacyl tRNA-synthetase (aaRS) native to the cell to incorporate UAAs into its proteome.[57] This method takes advantage of the promiscuity of the native aaRS; in some cases, the aaRS can charge UAAs that are sterically similar to a natural amino acid (NAA). This allows for the incorporation of

---

[54] For recent reviews, see: a) Chin, J. W. Expanding and Reprogramming the Genetic Code. *Nature* **2017**, *550*, 53–60. b) Young, D. D.; Schultz, P. G. Playing with the Molecules of Life. *ACS Chem. Biol.* **2018**, *13*, 854–870.

[55] a) Conibear, A. C.; Watson, E. E.; Payne, R. J.; Becker, C. F. W. Native Chemical Ligation in Protein Synthesis and Semi-Synthesis. *Chem. Soc. Rev.* **2018**, *47*, 9046–9068. b) Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J. C. M.; Melnyk, O. Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019**, *119*, 7328–7443.

[56] a) Quast, R. B.; Mrusek, D.; Hoffmeister, C.; Sonnabend, A.; Kubick, S. Cotranslational Incorporation of Non-Standard Amino Acids Using Cell-Free Protein Synthesis. *FEBS Lett.* **2015**, *589*, 1703–1712. b) Gregorio, N. E.; Levine, M. Z.; Oza, J. P. A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc.* **2019**, *2*, 24. c) Gao, W.; Cho, E.; Liu, Y.; Lu, Y. Advances and Challenges in Cell-Free Incorporation of Unnatural Amino Acids into Proteins. *Front. Pharmacol.* **2019**, *10*, 1–8.

[57] a) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. Residue-Specific Incorporation of Non-Canonical Amino Acids into Proteins: Recent Developments and Applications. *Curr. Opin. Chem. Biol.* **2010**, *14*, 774–780. b) Ngo, J. T.; Tirrell, D. A. Noncanonical Amino Acids in the Interrogation of Cellular Protein Synthesis. *Acc. Chem. Res.* **2011**, *44*, 677–685.

unnatural analogues that are isosteric to NAAs and can compete as a substrate for protein expression.

The simplest residue specific method for expressing UAA-containing proteins in *E. coli* is performed via selective pressure incorporation (SPI).[58] For the SPI method to work, an auxotrophic cell strain that is deficient in the biosynthetic machinery needed to make the target NAA is necessary along with control over the growth media. The cells are initially grown in NAA-containing media followed by a media exchange to media containing the UAA and lacking NAA. Induction of a target recombinant protein translation is done after the media switch to reduce the amount of NAA incorporation. Once the UAA is added, the aaRS incorporates it into the entire proteome in the cell in the positions the NAA residue normally occupies.

This approach is considered simple because it doesn't require the engineering of a tailored aaRS for the unnatural amino acid. However, despite its simplicity, it has a few disadvantages. The number of UAAs that can be incorporated using this method is limited to amino acids that are sterically similar to NAAs and are a substrate for the natural amino acid's aaRS. Additionally, once the media exchange occurs and initiation of protein synthesis is started, the UAA is incorporated into both the over-expressed recombinant protein and the entire proteome of the cell, which can cause issues with cell growth. Lastly, there is no way to select for a specific site in the protein as it will replace all occurrences of the natural amino acid with the UAA.

Other methods for residue specific incorporation have been developed for amino acids that cannot be incorporated via SPI using the endogenous tRNA synthetases. Mutant

[58] Budisa, N. Prolegomena to Future Experimental Efforts on Genetic Code Engineering by Expanding Its Amino Acid Repertoire. *Angew. Chem. Int. Ed.* **2004**, *43*, 6426–6463.

aminoacyl-tRNA-synthetases can be developed to direct the incorporation of UAAs that are not substrates for native synthetases.[32] Methods that have been used to develop and increase the selectivity of synthetases for UAAs include synthetase active site engineering[59], aaRS overexpression[60], and mutations of editing domains.[61]

While residue-specific methods target the replacement of one NAA throughout the proteome, the site-specific method targets one site in a target recombinant protein. UAAs can be incorporated in a site-specific fashion *via* genetic code expansion, which is accomplished through the use of an orthogonal aaRS/tRNA pair to insert a UAA *via* nonsense or frameshift codon suppression.[62] In this technique, the active site of an orthogonal aaRS must be evolved to select for a UAA without cross-reactivity with the 20 endogenous amino acids.

In comparison with the residue-specific method, the site-specific method allows the researcher to insert a UAA at one target site in a protein rather than at every occurrence of

---

[59] For examples, see: a) Datta, D.; Wang, P.; Carrico, I. S.; Mayo, S. L.; Tirrell, D. A. A Designed Phenylalanyl-tRNA Synthetase Variant Allows Efficient in Vivo Incorporation of Aryl Ketone Functionality into Proteins. *J. Am. Chem. Soc.* **2002**, *124*, 5652–5653. b) Sharma, N.; Furter, R.; Kast, P.; Tirrell, D. A. Efficient Introduction of Aryl Bromide Functionality into Proteins in Vivo. *FEBS Lett.* **2000**, *467*, 37–40.

[60] For an example, see: Kiick, K. L.; Van Hest, J. C. M.; Tirrell, D. A. Expanding the Scope of Protein Biosynthesis by Altering the Methionyl- tRNA Synthetase Activity of a Bacterial Expression Host. *Angew. Chem. Int. Ed.* **2000**, *39*, 2148–2152.

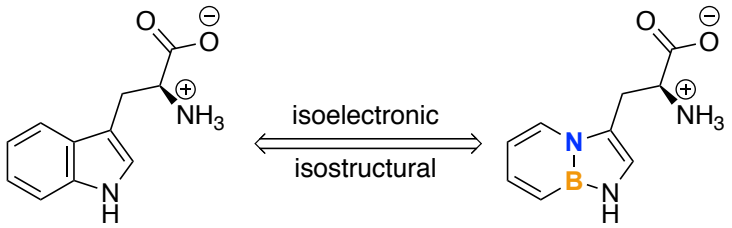
[61] For examples, see: a) Döring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; De Crécy-Lagard, V.; Schimmel, P.; Marlière, P. Enlarging the Amino Acid Set of Escherichia Coli by Infiltration of the Valine Coding Pathway. *Science*. **2001**, *292*, 501–504. b) Tang, Y.; Tirrell, D. A. Attenuation of the Editing Activity of the Escherichia Coli Leucyl-tRNA Synthetase Allows Incorporation of Novel Amino Acids into Proteins in Vivo. *Biochemistry* **2002**, *41*, 10635–10645. c) Tang, Y.; Wang, P.; Van Deventer, J. A.; Link, A. J.; Tirrell, D. A. Introduction of an Aliphatic Ketone into Recombinant Proteins in a Bacterial Strain That Overexpresses an Editing-Impaired Leucyl-tRNA Synthetase. *ChemBioChem* **2009**, *10*, 2188–2190.

[62] a) Cropp, T. A.; Schultz, P. G. An Expanding Genetic Code. *Trends Genet.* **2004**, *20*, 625–630. b) Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **2014**, *83*, 379–408. c) Wals, K.; Ovaa, H. Unnatural Amino Acid Incorporation in E. Coli: Current and Future Applications in the Design of Therapeutic Proteins. *Front. Chem.* **2014**, *2*, 1–12. d) Smolskaya, S.; Andreev, Y. A. Site-Specific Incorporation of Unnatural Amino Acids into Escherichia Coli Recombinant Protein: Methodology Development and Recent Achievement. *Biomolecules* **2019**, *9*, 255. e) Wang, K.; Schmied, W. H.; Chin, J. W. Reprogramming the Genetic Code: From Triplet to Quadruplet Codes. *Angew. Chem. Int. Ed.* **2012**, *51*, 2288–2297.

a NAA. It is a much more intensive method, however, since a specialized aaRS/tRNA pair must be evolved for the specific UAA. It is also difficult to incorporate more than one UAA, however, recent efforts have made this possible.[63]

1.3. BN-Tryptophan

Figure 1.18 BN-Isostere of Trytophan



Tryptophan 1.60 **BN-Tryptophan 1.61**

Due to the presence of tryptophan in biological studies and the gap in the current literature for application of the "fused" BN-indole **1.18**, we decided to synthesize the BN-analogue of tryptophan **1.61** (Figure 1.18). The addition of a BN-analogue of tryptophan to the UAA toolbox could offer an alternative approach to probing protein function and fluorescent properties while also demonstrating the utility of azaborine chemistry in a biological context. We believe this BN-analogue of tryptophan to be an ideal substrate for protein studies using SPI due to BN-indole's isosteric relationship with indole.

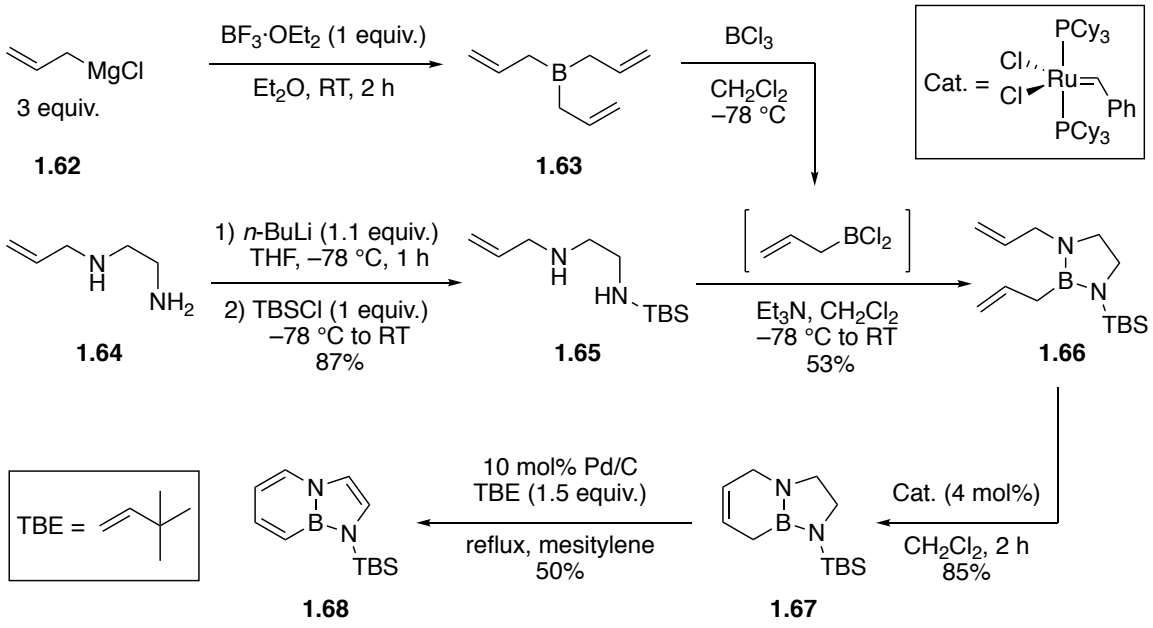
---

[63] For examples, see: a) Neumann, H.; Wang, K.; Davis, L.; Garcia-Alai, M.; Chin, J. W. Encoding Multiple Unnatural Amino Acids via Evolution of a Quadruplet-Decoding Ribosome. *Nature* **2010**, *464*, 441–444. b) Wan, W.; Huang, Y.; Wang, Z.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R. A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia Coli. *Angew. Chem. Int. Ed.* **2010**, *49*, 3211–3214. c) Wang, K.; Sachdeva, A.; Cox, D. J.; Wilf, N. W.; Lang, K.; Wallace, S.; Mehl, R. A.; Chin, J. W. Optimized Orthogonal Translation of Unnatural Amino Acids Enables Spontaneous Protein Double-Labelling and FRET. *Nat. Chem.* **2014**, *6*, 393–403. d) Zheng, Y.; Gilgenast, M. J.; Hauc, S.; Chatterjee, A. Capturing Post-Translational Modification-Triggered Protein-Protein Interactions Using Dual Noncanonical Amino Acid Mutagenesis. *ACS Chem. Biol.* **2018**, *13*, 1137–1141. e) Italia, J. S.; Addy, P. S.; Erickson, S. B.; Peeler, J. C.; Weerapana, E.; Chatterjee, A. Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* **2019**, *141*, 6204–6212. f) Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia Coli. *Biochemistry* **2013**, *52*, 1828–1837.

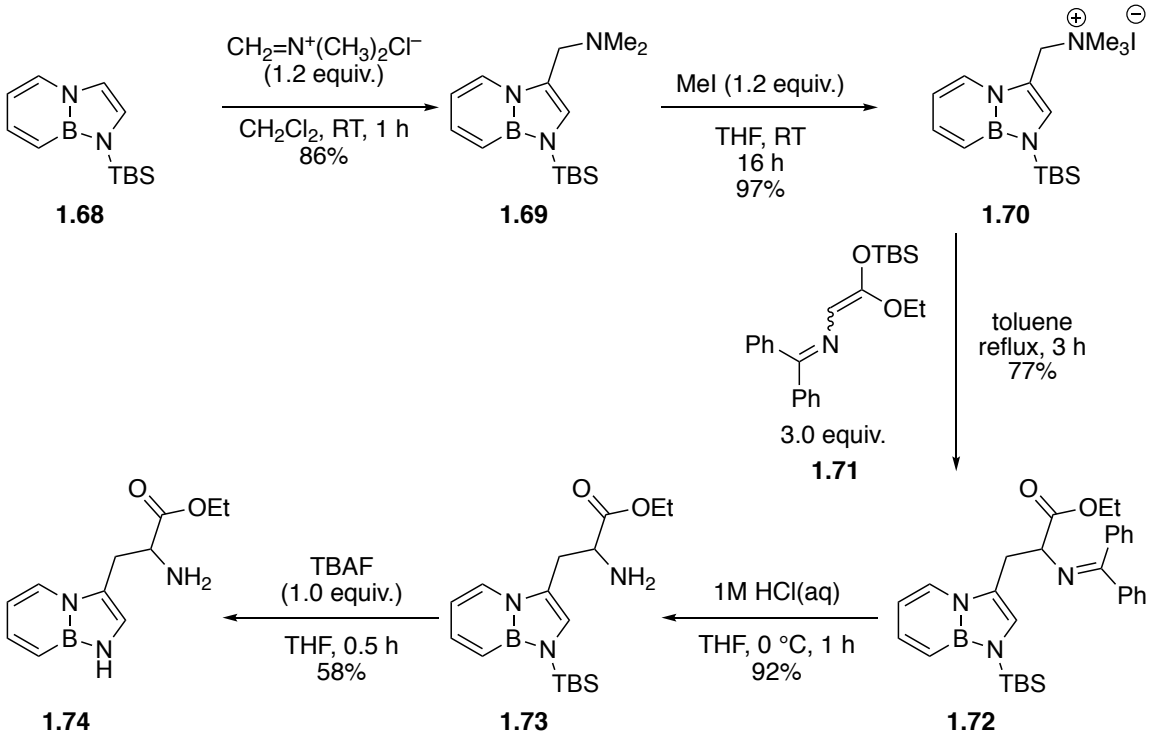
1.3.1. Synthesis and Chiral Resolution of BN-tryptophan

We synthesized the TBS-protected BN-indole **1.68** through a slightly modified version of the previously reported method from the Liu lab (Scheme 1.1).[22] In place of the allyltriphenyl tin,[22] we started with triallyl borane **1.63**, which we synthesized through a Grignard reaction of boron trifluoride diethyl etherate and *in situ* generated allyl magnesium chloride. The triallyl borane **1.63** was then reacted with boron trichloride to make allylboron dichloride *in situ*, followed by condensation with *N*-TBS-*N*-allyl ethylene diamine **1.65** (synthesized from *N*-allyl ethylene diamine **1.64**). The resulting diallyl species **1.66** was then subjected to ring-closing metathesis with Grubbs First Generation Catalyst. A modified aromatization procedure using mesitylene and 3,3-dimethyl-1-butene afforded the aromatic, TBS-protected BN-indole **1.68**.

Scheme 1.1 Synthesis of BN-Indole **1.68**

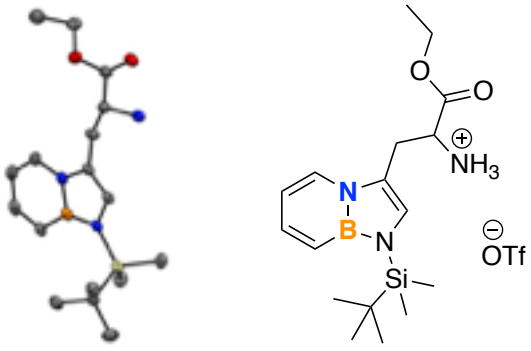


Scheme 1.2 Synthesis of BN-Trytophan Ester **1.74**



With the BN-indole in hand, we used the previously reported EAS method with dimethyliminium chloride (Scheme 1.2).[21] The EAS product **1.69** was then methylated with iodomethane and subsequently displaced with silyl-ketene-acetal **1.71**[64] at an elevated temperature. Subjection to aqueous acidic conditions removed the Schiff base protecting

Figure 1.19 Crystal Structure of BN-Tryptophan Ester Triflate Salt

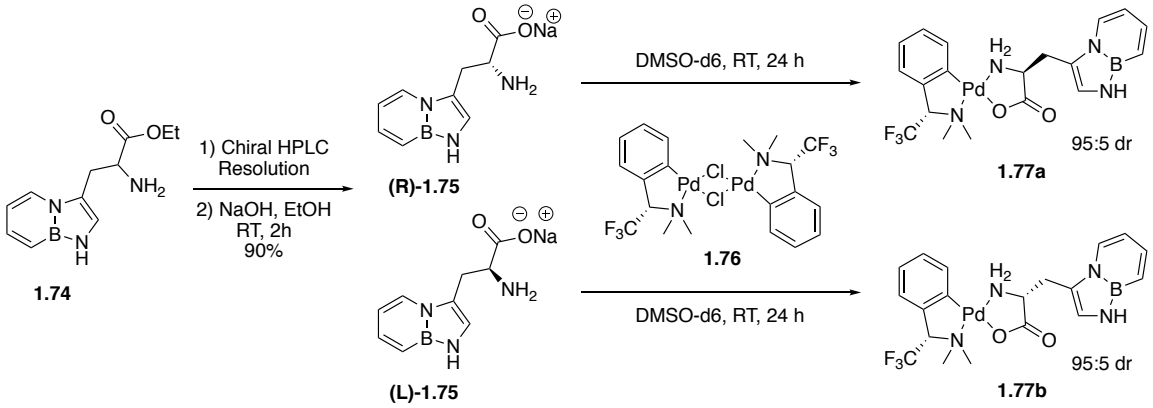


[64] Horikawa, M.; Busch-Petersen, J.; Corey, E. J. Enantioselective Synthesis of β-Hydroxy-α-Amino Acid Esters by Aldol Coupling Using a Chiral Quaternary Ammonium Salt as Catalyst. *Tetrahedron Lett.* **1999**, *40*, 3843–3846.

group.[65] Lastly, the silyl-protecting group was removed with tetrabutylammonium fluoride to afford the BN-tryptophan ethyl ester **1.74**. A crystal structure was also acquired for the triflate salt of ester **1.73** (Figure 1.19).

In order to directly compare properties with natural tryptophan, we needed to resolve the enantiomers. We accomplished this *via* chiral recycling HPLC. With the enantiomers of **1.74** in hand, we performed the final deprotection of the ester through basic hydrolysis to afford the parent BN-tryptophan amino acid salt **1.75** (Scheme 1.3). To confirm the HPLC separation was effective and the hydrolysis step did not racemize the amino acid stereocenter, we complexed it to a chiral palladium species **1.76**[66] and determined the enantiopurity by ^{1}H NMR. By NMR, we determined that the complexes **1.77a** and **1.77b** were in 95:5 diastereomeric ratio.

Scheme 1.3 Synthesis of BN-Tryptophan **1.75** and Chiral Analysis

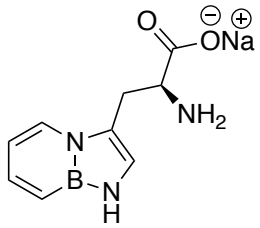


[65] Wartmann, T.; Lindel, T. L-Phototryptophan. *European J. Org. Chem.* **2013**, *9*, 1649–1652.

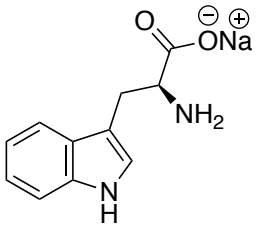
[66] Levrat, F.; Stoeckli-evans, H.; Engel, N. Enantiomeric Excess Determination of α-Amino Acids by ^{19}F NMR Spectroscopy of Their N,N-Dimethyl-(2,2,2-trifluoro-1-phenylethyl)amine-C,N)Palladium Complexes. *Tetrahedron: Asymmetry* **2002**, *13*, 2335–2344.

1.3.2. Characterization of BN-Tryptophan

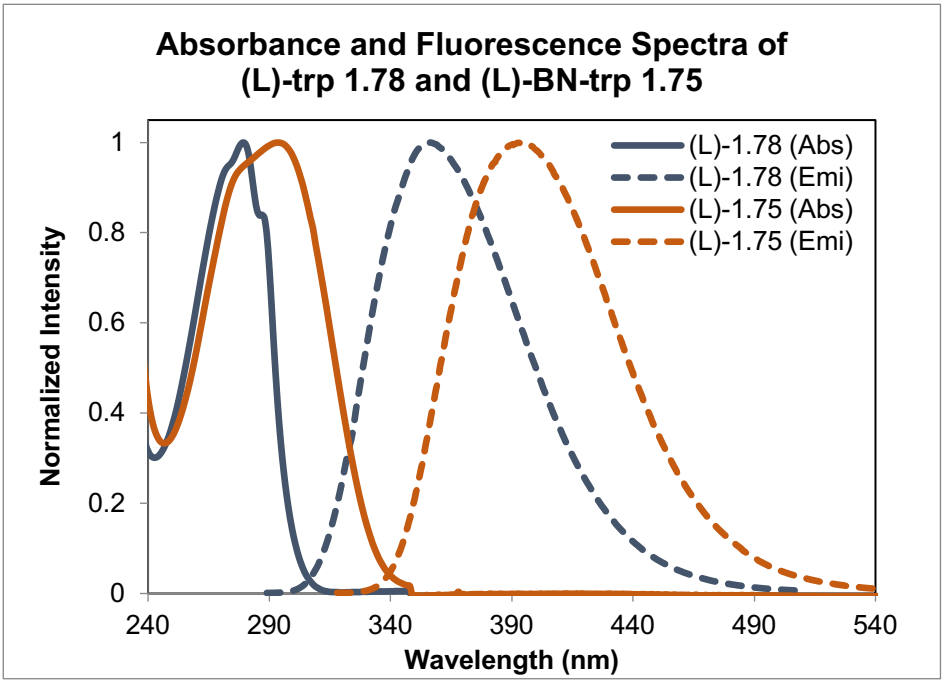
Figure 1.20 Spectroscopic Properties of (L)-Trp vs. (L)-BN-Trp



(L)-1.75



(L)-1.78



	λ_{abs}	λ_{emi}	Stokes Shift	Φ (H_2O, pH 10)
(L)-trp 1.78	279 nm	356 nm	7752 cm^{-1}	0.25 (0.31)[65]
(L)-BN-trp 1.75	294 nm	394 nm	8633 cm^{-1}	0.29

We determined the absorption and emission properties of the resulting amino acid sodium salts in comparison with tryptophan **(L)-1.78** (Figure 1.20). Tryptophan **(L)-1.78**'s

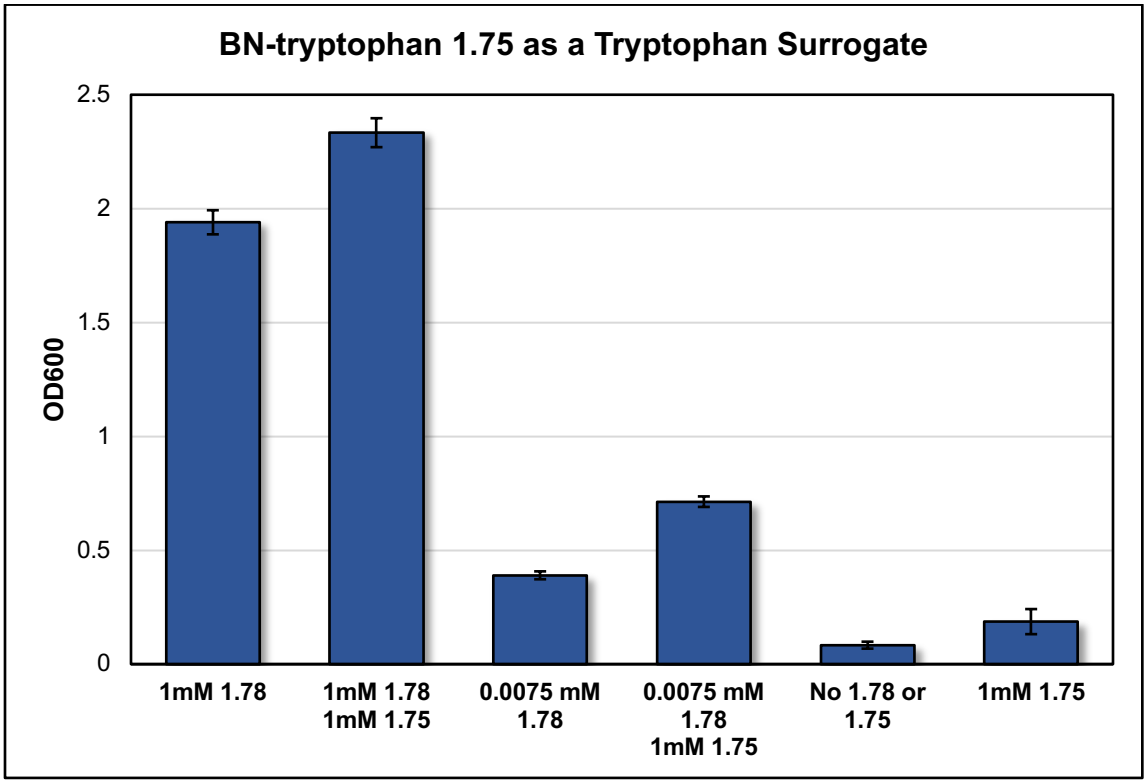
absorbance maximum was found at λ = 279 nm in water,[67] whereas BN-tryptophan **(L)-1.75**'s absorbance maximum is bathochromically shifted to λ = 294 nm. The fluorescence spectrum of (L)-BN-tryptophan shows an emission maximum of λ = 356 nm while L-tryptophan shows a maximum at λ = 394 nm.[67] Consequently, (L)-BN-tryptophan shows a larger Stokes shift of 8633 cm^{-1} compared to 7752 cm^{-1} of (L)-tryptophan. Both display a similar quantum yield of ~ 0.3.[67] The solvatochromic properties were explored as well and were consistent with previous studies (See Section 1.5.6).[68]

1.3.3. Incorporation of BN-Tryptophan into Green Fluorescent Proteins

With BN-tryptophan **1.75** in hand, we evaluated the possibility of its incorporation into proteins as a surrogate of L-tryptophan **1.78** using the SPI method with a tryptophan auxotrophic strain of *E. coli*.[52b] We demonstrated that BN-tryptophan does not inhibit cell growth in the presence of natural tryptophan, confirming its lack of general toxicity (Figure 1.21, columns 2 and 4). In a defined growth medium lacking tryptophan, addition of BN-tryptophan led to weak *E. coli* growth relative to a culture that did not receive any tryptophan analogue (Figure 1.21, column 6 vs. 5). However, addition of L-tryptophan to the same culture resulted in significantly higher growth. This indicates that BN-tryptophan is accepted as a substrate for the endogenous tryptophanyl-tRNA synthetase, but proteome-wide replacement of L-tryptophan with BN-tryptophan is not well-tolerated by *E. coli*.

[67] Szabo, A. G.; Rayner, D. M. Fluorescence Decay of Tryptophan Conformers in Aqueous Solution. *J. Am. Chem. Soc.* **1980**, *102*, 554–563.
[68] Saif, M.; Widom, J. R.; Xu, S.; Abbey, E. R.; Liu, S.-Y.; Marcus, A. H. Electric Dipole Transition Moments and Solvent-Dependent Interactions of Fluorescent Boron-Nitrogen Substituted Indole Derivatives. *J. Phys. Chem. B* **2015**, *119*, 7985–7993.

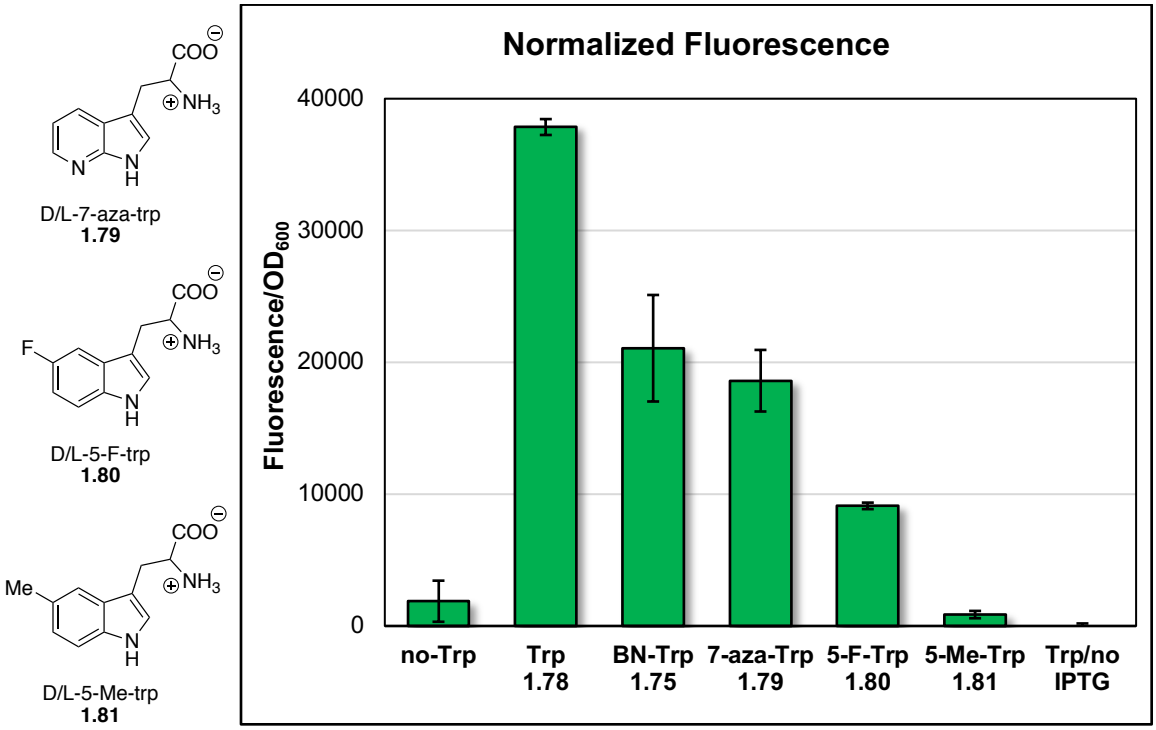
Figure 1.21 *E. coli* Cell Viability with BN-tryptophan **1.75**



To further our studies, we incorporated our UAA along with three selected tryptophan analogues, which are known substrates for the endogenous tryptophanyl-tRNA synthetase, into superfolder green fluorescent protein (sfGFP).[69] In this protein, there is a single tryptophan residue which precedes the chromophore. Expression of the full-length fluorescent sfGFP is reliant on successful decoding of the TGG codon, which codes for tryptophan. Indeed, this is what we see; the similar levels of fluorescence normalized with respect to optical density supports the notion that BN-tryptophan was incorporated into the sfGFP (Figure 1.22, columns 2 and 3). We also see similar or improved incorporation efficiency of BN-trp relative to 7-aza-tryptophan **1.79**, 5-F-tryptophan **1.80**, and 5-Me-tryptophan **1.81**, which are established surrogates of L-tryptophan (Figure 1.22).

[69] Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.

Figure 1.22 Fluorescence/OD_{600} Plot – A Comparison of Tryptophan Analogues



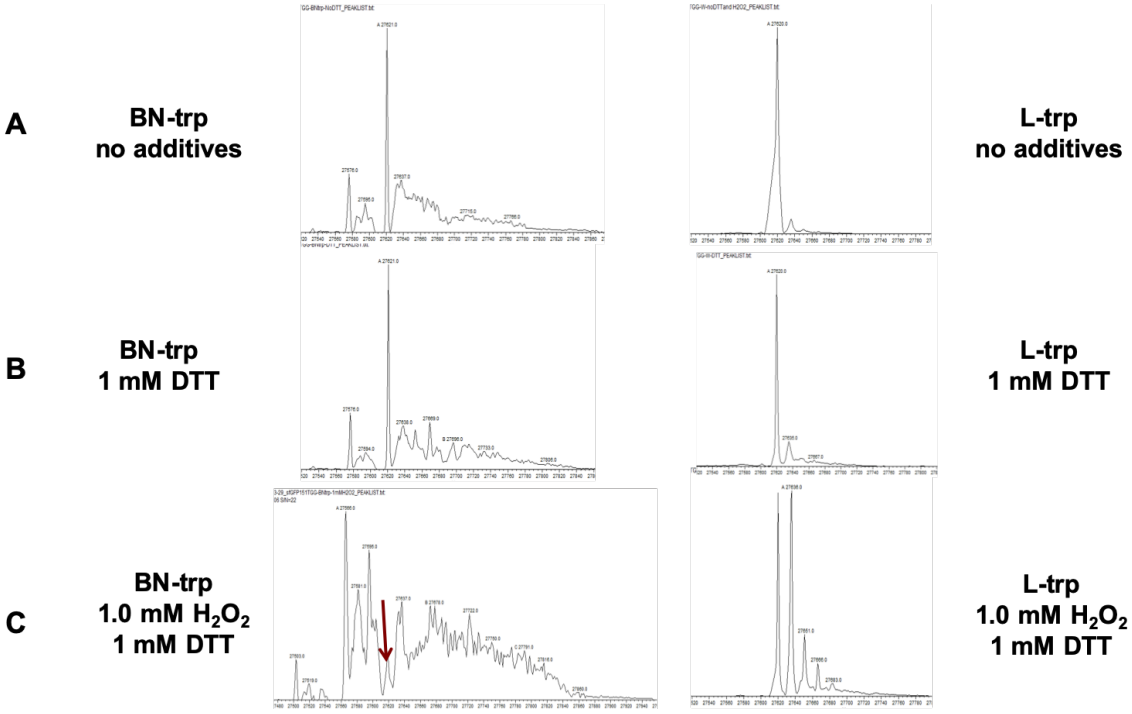
1.3.4. Reactivity of BN-Tryptophan

We also incorporated BN-tryptophan **1.75** into a sfGFP mutant which harbors an additional tryptophan codon (sfGFP-151-TGG)[70] at a surface exposed site and isolated the protein by immobilized metal-ion chromatography using a C-terminal poly-histidine tag. Analysis by LC/MS shows the expected mass (Figure 1.23). To further confirm the presence of the BN-tryptophan **1.75** in this protein, which has essentially the same mass as natural tryptophan **1.78**, we took advantage of its unique sensitivity to the mild oxidant hydrogen peroxide. sfGFP incorporating either tryptophan or BN-tryptophan was briefly subjected to 1 mM hydrogen peroxide, quenched by dithiothreitol (DTT), and analyzed by full-protein MS analysis. It was found that, as shown in Figure 1.23, the protein containing

[70] J. S. Italia, P. S. Addy, C. J. J. Wrobel, L. A. Crawford, M. J. Lajoie, Y. Zheng, A. Chatterjee *Nat. Chem. Biol.* **2017**, *13*, 446–450.

tryptophan underwent partial oxidation, which is expected. In stark contrast, the identical treatment led to near-complete loss of the original peak for sfGFP-BN-trp. Since these two otherwise identical proteins only differ by the presence of BN-tryptophan, we can conclude that BN-tryptophan displays different reactivity from tryptophan, a characteristic that can potentially be utilized in future studies.

Figure 1.23 Reactivity of BN-Tryptophan in a Fluorescent Protein

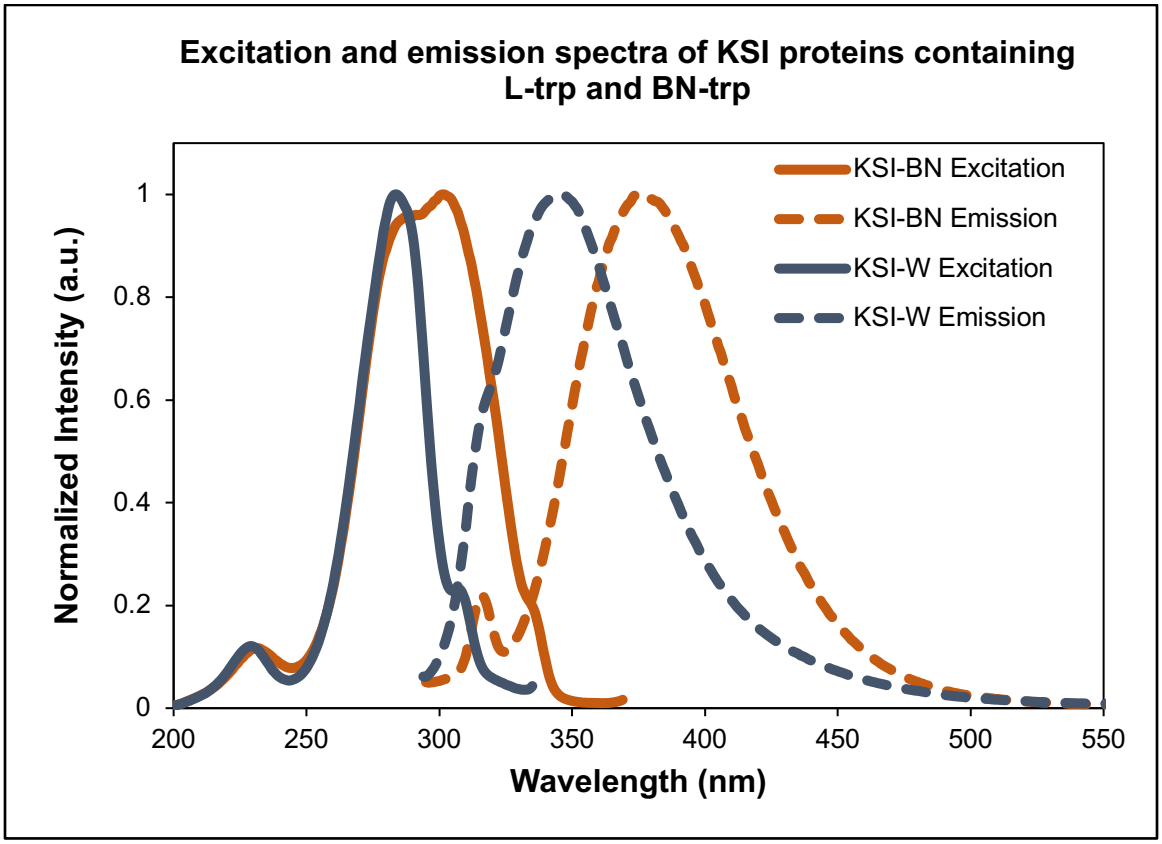


1.3.5. Incorporation of BN-tryptophan into Ketosteroid Isomerase

Since tryptophan serves as the primary source of protein fluorescence, the incorporation of the BN-tryptophan should result in a red shift in the excitation and emission spectra when compared with the wild-type protein. To probe this property, we incorporated BN-tryptophan **1.75** into the non-fluorescent protein, ketosteroid isomerase (KSI), which is a homo-dimeric protein that houses two tryptophan residues in each

monomer.[71] As expected, the fluorescence of the BN-tryptophan-containing protein (KSI-BN) was red-shifted with respect to the L-tryptophan-containing protein (KSI-W). The KSI-BN and KSI-W fluorescence maxima were found to be 372 and 342 nm, respectively.[71] In addition, we are able to distinguish the tyrosine fluorescence peak (λ_{max} = 316 nm) from the BN-trp peak, while in the L-trp case, the peaks are not resolved (Figure 1.24).[72]

Figure 1.24 Spectroscopic Properties of a BN-Tryptophan-containing Protein



[71] Wang, S.-F.; Kawahara, F. S.; Talalay, P. The Mechanism of the Δ^5-3-Ketosteroid Isomerase Reaction: Absorption and Fluorescence Spectra of Complexes. *J. Biol. Chem.* **1963**, *238*, 576–585.
[72] Duggan, D. E.; Udenfriend, S. The Spectrophotofluorometric Determination of Tryptophan in Plasma and of Trytptophan and Tyrosine in Protein Hydrolysates. *J. Biol. Chem.* **1956**, *223*, 313–319.

1.4. Conclusions

In summary, we synthesized a BN-analogue of the canonical amino acid, tryptophan, and characterized its spectroscopic properties with respect to its natural tryptophan counterpart. We demonstrated through selective pressure incorporation that BN-tryptophan can be incorporated into proteins *in vivo* and that three different proteins (sfGFP, sfGFP-151-TGG, and KSI) containing BN-tryptophan can be isolated. This work shows that the natural tryptophanyl-tRNA synthetase can recognize an azaborine containing amino acid, further solidifying the potential use of BN/CC isosterism of arenes in a biological context.

1.5. Experimental Section

1.5.1. General Considerations

All oxygen- and moisture-sensitive manipulations were carried out under N_2 using either standard Schlenk techniques or a nitrogen-filled glovebox. THF, Et_2O, CH_2Cl_2, toluene, and pentane were dried with a solvent purification system consisting of columns of molecular sieves under argon. Silica gel (230-400 mesh) was dried for 12 hours at 180 °C under high vacuum. Flash chromatography was performed with this silica gel under an inert atmosphere. All other chemicals and solvents were purchased and used as received.

NMR spectra were recorded on a Varian VNMRS 600 MHz, VNMRS 500 MHz, INOVA 500 MHz, or VNMRS 400 MHz spectrometer. Deuterated solvents were purchased from Cambridge Isotope Labs. ^{11}B NMR spectra were externally referenced to $BF_3 \bullet Et_2O$ (δ 0). All NMR chemical shifts are reported in ppm relative to residual solvent for ^{1}H and ^{13}C NMR.

Infrared spectroscopy was performed on a Bruker ALPHA-Platinum FT-IR Spectrometer with ATR-sampling module.

High-resolution mass spectra were collected by Marek Domin on a JEOL AccuTOF instrument (JEOL USA, Peabody, MA), equipped with a DART ion source (IonSense, Inc., Danvers, MA) in positive ion mode at the Boston College Center for Mass Spectrometry.

UV-vis absorption spectra were collected on an Agilent Cary 100 UV-Vis spectrometer. Emission spectra were collected on a Photon Technology International spectrometer. Quantum yield (Φ) was determined using a PTI K-Sphere "Petite" integrating sphere.

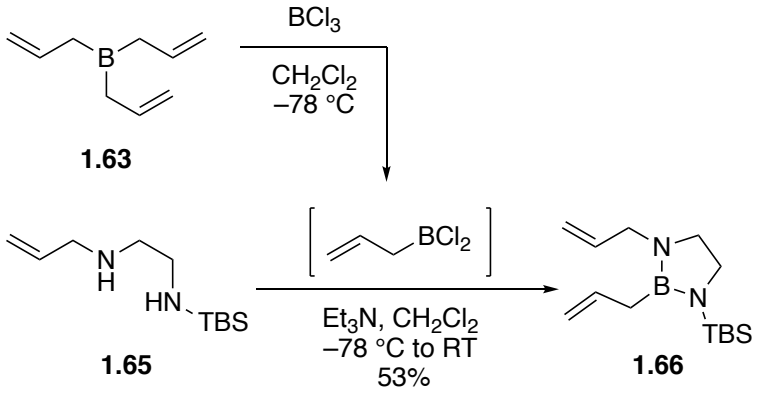
Analytical HPLC spectra were recorded on an Agilent 1200 Infinity Series HPLC equipped with a DAD detector. Recycling HPLC was performed on a Japan Analytical Co., Ltd. Next Recycling Preparative HPLC with UV/vis detector.

Specific rotations measured on a Rudolph Research Analytical Autopol IV.

Circular dichroism spectra were recorded on an Aviv Biomedical Inc. Model 420 Circular Dichroism Spectrometer.

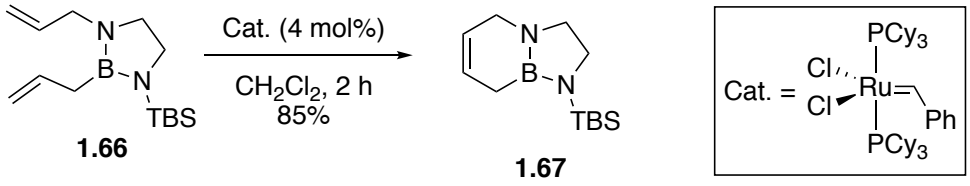
1.5.2. Synthesis of BN-tryptophan

Synthesis of Diallyl species **1.66:**



In a glove box, to a 1L round-bottom flask (RBF) containing a stir bar was added triallylborane **1.63** (6.71 g, 50.0 mmol) and CH_2Cl_2 (300 mL). The stirred solution was cooled to –78 °C in a dry ice/acetone bath under an atmosphere of nitrogen, and a solution of borontrichloride in hexanes (1.0 M, 100 mL) was added via cannula. The reaction was allowed to stir for 4 h at –78 °C, and then a solution of *N*-allyl-*N*'-TBS-ethylenediamine **1.65**[22] (32.5 g, 150 mmol) in CH_2Cl_2 (100 mL) was added dropwise via cannula at –78 °C. Then, triethylamine (18.5 mL, 132 mmol) was added dropwise via cannula, also at –78 °C. After addition, the reaction mixture was allowed to slowly warm to 25 °C for 16 h. At the conclusion of the reaction, the solvent was removed under reduced pressure and then transferred to a nitrogen-filled glove box. Pentane (150 mL) was added and a white precipitate formed. The mixture was passed through a glass frit in a glove box with copious washing using pentane (~100mL). The filtrate was concentrated under reduced pressure. The oily crude residue was purified via vacuum distillation to afford a colorless cloudy liquid (21.2 g, 52.9%). Product decomposes when exposed to water and/or oxygen. Spectra of isolated compound matched previously published records.[22]

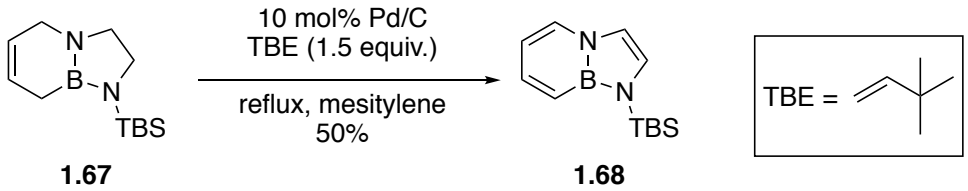
Synthesis of Ring-closed product **1.67:**



1.66 **1.67**

In a glove box, a 1L RBF equipped with a magnetic stir bar was charged with a solution of diallyl BN_2-adduct **1.66** (21.3 g, 80.7 mmol) in CH_2Cl_2 (500 mL). Grubbs Gen. 1 catalyst (bis(tricyclohexylphosphine) benzylidine ruthenium(IV) chloride) (2.66 g, 3.23 mmol) was added portion wise to the stirring solution. The reaction was stirred for 2 h under a nitrogen atmosphere at room temperature and was monitored by 1H NMR. At the conclusion of the reaction, the solvent was removed under reduced pressure. The crude material was purified by vacuum distillation (b.p. 63 °C, 120 mTorr) to afford the clear colorless liquid **1.67** (15.01 g, 86%). Product decomposes when exposed to water and/or oxygen. Spectra of isolated compound matched previously published records.[22]

Synthesis of BN-Indole **1.68:**

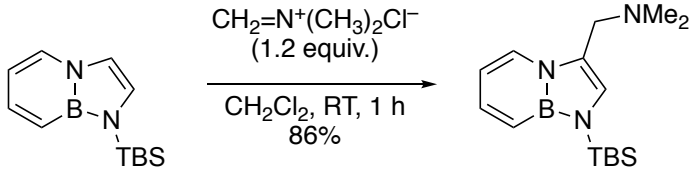


1.67 **1.68**

In a glove box, a 500 mL RBF containing a stir bar was charged with palladium, 10% on carbon (2.24 g, 21.2 mmol, Acros), ring-closed product **1.67** (5 g, 21.2 mmol), 3,3-dimethylbut-1-ene (1.96 g, 23.3 mmol, 3 mL) and mesitylene (90 mL). The reaction was brought to reflux under a nitrogen atmosphere. After 16 hours, the reaction was cooled to

room temperature. The suspension was filtered through a glass frit, washed with THF, and solvent was removed under vacuum distillation. The crude oil was passed through a silica plug in the glove box with DCM as eluent. Resulting oil was fractionally distilled under reduced pressure (b. p. 76 °C at 300 mTorr) to isolate a clear oil, TBS-BN-indole **1.68** (2.75 g, 11.8 mmol, 56.0% yield). Spectra of isolated compound matched previously published records.[22]

Synthesis of TBS-BN-gramine **1.69**:



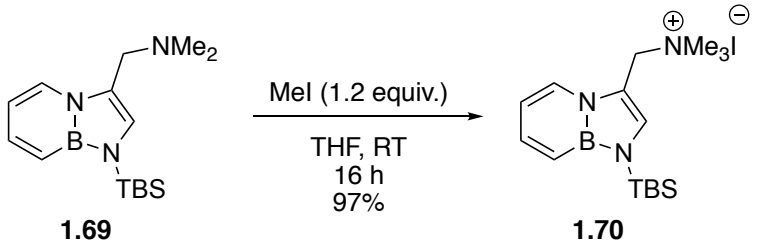
1.68 **1.69**

In a glove box, a 500 mL RBF equipped with a stir bar was charged with TBS-BN-indole **1.68** (2.0 g, 8.6 mmol) and *N,N′*-dimethylmethylene-iminium chloride (967 mg, 10.3 mmol) and dissolved in CH_2Cl_2 (100 mL). The reaction was allowed to stir for 3 hours. Solvent was removed under reduced pressure and, in a glove box, the resulting yellow solids were suspended in 15 mL triethylamine and 15 mL diethyl ether and passed through a glass frit. After the solvents were removed under reduced pressure, the product was purified by silica gel chromatography under a nitrogen atmosphere with 15% triethylamine in diethyl ether as eluent. TBS-BN-gramine **1.69** was concentrated under reduced pressure to yield an opaque, light yellow oil (2.15 g, 7.43 mmol, 86.3% yield).

^{1}H NMR (500 MHz, CD_2Cl_2) δ 8.01 (d, J = 6.5 Hz, 1H), 7.57 (dd, J = 11.4, 6.5 Hz, 1H), 6.76 (d, J = 11.5 Hz, 1H), 6.62 (s, 1H), 6.38 (t, J = 7.0 Hz, 1H), 3.52 (s, 2H), 2.22 (s, 6H), 0.91 (s, 9H), 0.47 (s, 6H). ^{13}C NMR (126 MHz, CD_2Cl_2) δ 138.69, 128.51, 123.57,

123.56, 123.40, 108.47, 55.08, 45.38, 26.51, 19.16, -4.13 (B-C signal not observed). ^{11}B NMR (160 MHz, CD_2Cl_2) δ 26.5. IR (ATR) 2950, 2928, 2855, 2813, 2763, 1618, 1587, 1499, 1426, 1347, 1310, 1173, 1162, 1014, 947, 834, 806, 781, 728, 698, 678, 429 cm^{-1}. HRMS (DART+) calcd. for $C_{15}H_{28}B_1N_3Si_1$ $[M+H]^+$ 289.21455, found 289.21513.

Synthesis of TBS-BN-gramine methiodide salt **1.70**:



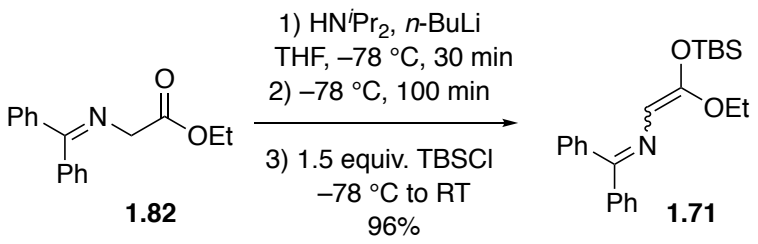
1.69 **1.70**

In a glove box, to a 250 mL RBF equipped with a stir bar was added TBS-BN-gramine **1.69** (1.70 g, 5.88 mmol) and THF (75 mL). The flask was cooled to 0 °C in an ice bath under a nitrogen atmosphere for 15 minutes and iodomethane (1.00 g, 7.05 mmol, 439 μL) was added dropwise via syringe. The reaction was allowed to warm to 25 °C for 16 hours. The solids were allowed to settle in the bottom of the flask and the supernatant was removed via cannula under nitrogen pressure. When most of the solvent was removed, the solids were resuspended in CH_2Cl_2 and precipitated out with THF and the supernatant was removed by cannula. The residual solvent was removed under reduced pressure. TBS-BN-gramine methiodide salt **1.70** was obtained as a white, powdery solid (2.45 g, 5.68 mmol, 96.7% yield). Product decomposes when exposed to water and/or oxygen.

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.34 (d, J = 6.7 Hz, 1H), 7.57 (dd, J = 11.5, 6.5 Hz, 1H), 7.28 (s, 1H), 6.76 (d, J = 11.5 Hz, 1H), 6.50 (t, J = 6.6 Hz, 1H), 4.83 (s, 2H), 3.06 (s, 9H), 0.86 (s, 9H), 0.47 (s, 6H). ^{13}C NMR (151 MHz, dmso) δ 138.77, 130.74, 126.81, 114.32, 108.88, 58.26, 51.32, 25.72, 17.99, -4.82 (B-C signal not observed). ^{11}B NMR (160

MHz, DMSO-d_6) δ 26.2. IR (ATR) 3000, 2950, 2928, 2854, 1605, 1578, 1500, 1471, 1428, 1364, 1324, 1296, 1252, 1154, 1119, 999, 973, 940, 838, 807, 787, 728, 700,592,430cm^{-1}.

Synthesis of silyl-ketene-acetal **1.71**:

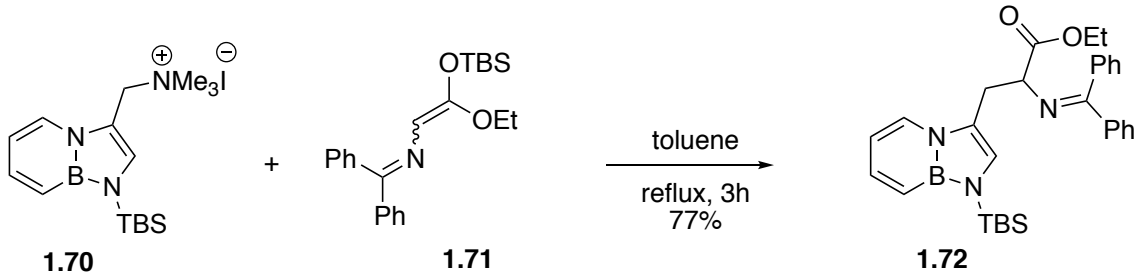


To a 1 L RBF equipped with a stir bar, diisopropyl amine (3.15 g, 31.1 mmol, 4.39 mL) and THF (300 mL) were added and cooled to –78 °C in a dry ice/acetone bath under an atmosphere of nitrogen. A solution of *n*-butyllithium (2.5 M in hexane, 28.7 mmol, 11.5 mL) was added dropwise via syringe and the reaction was allowed to stir for 30 min. *N*-(diphenylmethylene)glycine ethyl ester **1.82** (6.40 g, 24.0 mmol) was dissolved in THF (10 mL) and added to the reaction mixture dropwise via syringe. The resulting deep red solution was allowed to stir for 100 min at –78 °C, after which time a solution of *tert*-butyldimethylchlorosilane (5.05 g, 33.5 mmol) in THF (30 mL) was added dropwise via cannula. The reaction was allowed to stir and attain room temperature for 16 hours. Solvent was removed under reduced pressure and the crude mixture was suspended in hexane and passed through a glass frit. The filtrate was concentrated in vacuo. The desired product **1.71** was isolated as a red viscous oil (8.8 g, 96% yield). The product was used without further purification.

^{1}H NMR (500 MHz, CD_2Cl_2) δ 7.64 – 7.59 (m, 1H), 7.48 (ddd, J = 7.8, 6.6, 2.9 Hz, 2H), 7.42 (dt, J = 8.5, 2.0 Hz, 2H), 7.29 (dd, J = 4.5, 2.3 Hz, 3H), 7.22 – 7.17 (m, 2H), 5.62 (d, J = 9.9 Hz, 1H), 4.45 (q, J = 7.1 Hz, 1H), 3.69 (q, J = 7.0 Hz, 2H), 1.40 (t, J = 7.1 Hz,

1H), 1.21 (t, J = 7.0 Hz, 2H), 1.05 (s, 6H), 0.89 (s, 3H), 0.31 (s, 4H), 0.09 (s, 2H). ^{13}C NMR (126 MHz, CD_2Cl_2) δ 158.65, 155.83, 141.24, 141.07, 137.87, 137.66, 129.35, 129.30, 129.17, 129.15, 128.85, 128.63, 128.50, 128.40, 128.18, 127.97, 101.20, 98.19, 66.05, 64.06, 54.00, 26.10, 25.95, 18.95, 18.62, 15.86, 14.85, -3.90, -4.33. IR (ATR) 3059, 2953, 2958, 2884, 2855, 1624, 1470, 1444, 1376, 1315, 1294, 1240, 1070, 1017, 941, 826, 782, 695, 466 cm^{-1}. HRMS (DART+) calcd. for $C_{23}H_{32}N_1O_2Si_1$ 382.2202, found 382.2203.

Synthesis of protected BN-tryptophan **1.72**:



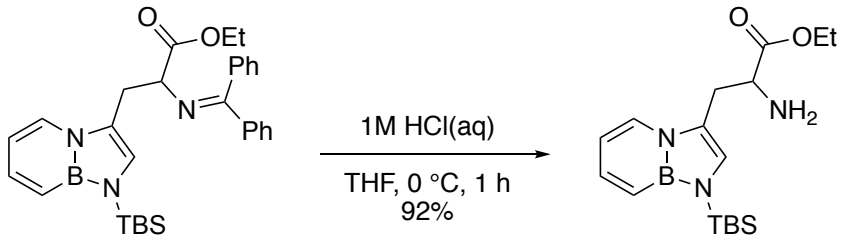
1.70 **1.71** **1.72**

In a glove box, a 1 L RBF equipped with a stir bar was charged with TBS-BN-gramine methiodide salt **1.70** (2.65 g, 6.20 mmol), silylketeneacetal **1.71** (7.10 g, 18.6 mmol), and toluene (400 mL). The reaction was brought to reflux under nitrogen and allowed to heat at 140 °C. After 3 hours, the reaction was cooled to room temperature and concentrated under reduced pressure. The product was purified by silica gel chromatography with a gradient of ether in pentane (starting at 10% and then 40% ether) in the glove box to afford an orange goo (2.23 g, 76.9 %). R_f (2:3 ether:pentane) = 0.4

^{1}H NMR (500 MHz, CD_2Cl_2) δ 7.60 – 7.55 (m, 2H), 7.51 (dd, J = 11.5, 6.5 Hz, 1H), 7.42 – 7.25 (m, 5H), 7.15 (t, J = 7.5 Hz, 2H), 6.76 – 6.72 (m, 1H), 6.58 (s, 2H), 6.51 (s, 1H), 6.18 (t, J = 6.5 Hz, 1H), 4.31 (ddd, J = 9.1, 4.2, 1.0 Hz, 1H), 4.19 (qdd, J = 7.1, 4.5, 1.0 Hz, 2H), 3.40 (dd, J = 14.9, 4.2 Hz, 1H), 3.26 (dd, J = 15.0, 9.1 Hz, 1H), 1.27 (t, J

= 7.1 Hz, 3H), 0.81 (s, 9H), 0.37 (d, J = 15.3 Hz, 6H). ^{13}C NMR (126 MHz, CD_2Cl_2) δ 172.08, 171.38, 139.81, 138.40, 136.30, 130.87, 129.27, 128.76, 128.48, 127.85, 126.55, 123.15, 121.78, 108.69, 64.51, 61.61, 29.25, 26.40, 18.99, 14.60, -4.10 (B-C signal not observed). ^{11}B NMR (160 MHz, CD_2Cl_2) δ 25.75. IR (ATR) 3063, 2953, 2927, 2855, 1735, 1617, 1586, 1499, 1470, 1444, 1427, 1314, 1288, 1252, 1175, 1137, 1118, 1030, 945, 834, 806, 780, 728, 694, 428 cm^{-1}. HRMS (DART+) calcd. for $C_{30}H_{39}B_1N_3O_2Si_1$ $[M+H]^+$ 512.2905, found 512.2921.

Synthesis of TBS-BN-tryptophan ethyl ester **1.73**:

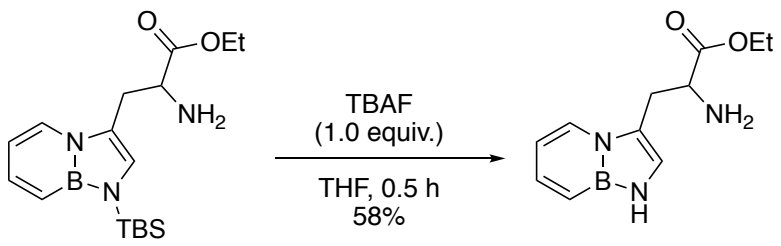


1.72 **1.73**

To a 250 mL RBF equipped with a stir bar was added a solution of **1.72** (1.4 g, 2.7 mmol) in THF (50 mL) which was cooled to 0 °C under an atmosphere of nitrogen in an ice bath. An aqueous solution of HCl (20 mL, 1M) was added dropwise via syringe and the solution was stirred at 0 °C for 30 minutes. The solution was diluted with 100 mL CH_2Cl_2 and washed thrice with 200 mL saturated sodium bicarbonate. The organic layer was separated and the aqueous layer was extracted twice more with 100 mL CH_2Cl_2. The combined organic layer was washed with brine and dried over $MgSO_4$. The solvent was removed under reduced pressure. The crude reaction mixture was purified by silica gel column, first eluted with CH_2Cl_2 and then 20% Et_3N in CH_2Cl_2, to yield **1.73** (0.867 g, 92%) as a brown oil. R_f (100 % ether) = 0.3

^{1}H NMR (500 MHz, CD_2Cl_2) δ 7.81 (d, J = 6.6 Hz, 1H), 7.57 (dd, J = 11.5, 6.5 Hz, 1H), 6.77 (d, J = 12.4 Hz, 1H), 6.60 (s, 1H), 6.41 (t, J = 6.6 Hz, 1H), 4.17 – 4.09 (m, 2H), 3.77 (dd, J = 7.6, 5.5 Hz, 1H), 3.22 (dd, J = 15.1, 5.5 Hz, 1H), 2.97 (dd, J = 15.1, 7.7 Hz, 1H), 1.52 (s, 2H), 1.22 (t, J = 7.1 Hz, 3H), 0.90 (s, 9H), 0.46 (s, 6H). ^{13}C NMR (126 MHz, CD_2Cl_2) δ 175.53, 138.55, 129.55, 126.76, 122.67, 121.87, 108.88, 61.46, 31.34, 26.51, 19.16, 14.54, -4.10 (B-C signal not observed). ^{11}B NMR (160 MHz, CD_2Cl_2) δ 26.0. IR (ATR) 2953, 2928, 2856, 1733, 1615, 1586, 1450, 1470, 1443, 1427, 1309, 1283, 1251, 1184, 1139, 1119, 1027, 945, 835, 820, 807, 782, 729, 699, 677, 428 cm^{-1}. HRMS (DART+) calcd. for $C_{17}H_{31}B_1N_3O_2Si_1$ $[M+H]^+$ 348.22786, found 348.22637.

Synthesis of BN-tryptophan ethyl ester **1.74**:

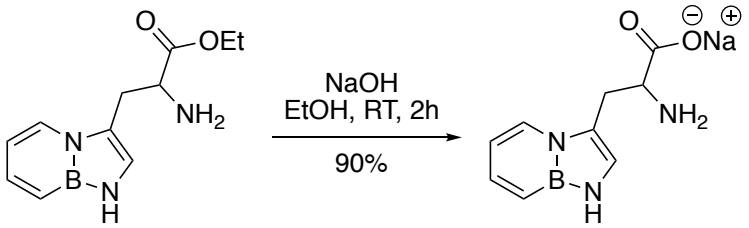


1.73 **1.74**

A 250 mL RBF was charged with **1.73** (473 mg, 1.36 mmol), THF (50 mL), and a stir bar and put under a nitrogen atmosphere. While stirring, tetrabutylammonium fluoride (1 M in THF, 1.36 mL) was added dropwise via syringe. After 30 minutes, the solvent was removed under reduced pressure. The resulting brown oil was suspended in 150 mL toluene. The organic layer was washed four times with 200 mL water followed by one wash with brine. The suspension was passed through filter paper and the solvent was removed under reduced pressure to afford compound **1.74** as a brown oil (185 mg, 58%).

^{1}H NMR (500 MHz, CD_2Cl_2) δ 7.81 (d, J = 6.6 Hz, 1H), 7.57 (dd, J = 11.4, 6.5 Hz, 1H), 6.82 – 6.50 (m, 3H), 6.41 (td, J = 6.6, 1.1 Hz, 1H), 4.15 (q, J = 7.1 Hz, 2H), 3.76 (dd, J = 7.9, 5.2 Hz, 1H), 3.27 – 3.19 (m, 1H), 2.96 (dd, J = 15.1, 7.9 Hz, 1H), 1.86 (s, 2H), 1.23 (t, J = 7.1 Hz, 3H). ^{13}C NMR (126 MHz, CD_2Cl_2) δ 175.10, 138.12, 126.03, 118.62, 118.15, 108.45, 61.06, 53.35, 30.71, 14.05 (B-C signal not observed). ^{11}B NMR (160 MHz, CD_2Cl_2) δ 23.44. IR (ATR) 3360, 2978, 2933, 1725, 1620, 1584, 1497, 1447, 1372, 1347, 1306, 1264, 1188, 1095, 1077, 1023, 857, 802, 730, 694, 586, 549, 519 cm^{-1}. HRMS (DART+) calcd. for $C_{11}H_{17}B_1N_3O_2Si_1$ $[M+H]^+$ 234.1414, found 234.1405.

Synthesis of BN-tryptophan **1.75**:



1.74 **1.75**

In a glove box, to a 20 mL vial was added a stir bar, **1.74** (284 mg, 1.22 mmol), sodium hydroxide (97.6 mg, 2.44 mmol), and ethanol (3.00 mL) and stirred for 2 hours. Most of the ethanol was removed under reduced pressure. 1 mL of ether was added to the solution, followed by 2 mL pentane, after which a brown and white solid precipitated. The solid was triturated twice in pentane and ether, solution removed by pipette in between triturations. BN-tryptophan **1.75** was obtained as a powdery tan solid (247 mg, 89.2%).

^{1}H NMR (500 MHz, DMSO-d_6) δ 8.34 (s, 1H), 7.91 (d, J = 5.9 Hz, 1H), 7.42 (dd, J = 10.8, 6.5 Hz, 1H), 6.73 (s, 1H), 6.61 (d, J = 11.3 Hz, 1H), 6.30 (t, J = 5.5 Hz, 1H), 3.24 (d, J = 13.4 Hz, 2H), 2.66 – 2.53 (m, 1H), 1.49 (s, 2H). ^{13}C NMR (126 MHz, DMSO-d_6) δ

178.81, 136.84, 126.46, 120.40, 117.85, 107.47, 54.37, 32.02 (B-C signal not observed). ^{11}B NMR (160 MHz, DMSO-d_6) δ 23.22. IR (ATR) 3445, 3370, 3350, 1561, 1495, 1467, 1449, 1406, 1346, 1326, 1306, 1074, 1051, 927, 878, 860, 728, 692, 608, 557, 541 cm^{-1}. HRMS (DART+) calcd. for $C_9H_{13}B_1N_3O_2$ 206.11008, found 206.11063.

1.5.3. Chiral Resolution of BN-tryptophan Ester

Enantiomers of ester **1.74** were separated by chiral recycling HPLC (Chiralpak AD-H, 25°C, 78:19.5:2.0:0.5 Hexane:Isopropanol:Methanol:*n*-butylamine, 5 mL/min). The product was cycled through 4-6 times before collection of each peak.

Collected fractions were analyzed by chiral analytical HPLC (Chiralpak AD-H, 25 °C, 78:19.5:2.0:0.5 Hexane:Isopropanol: Methanol:*n*-butylamine, 0.5 mL/min).

Figure 1.25: HPLC Trace of Racemic BN-tryptophan ester **1.74**.

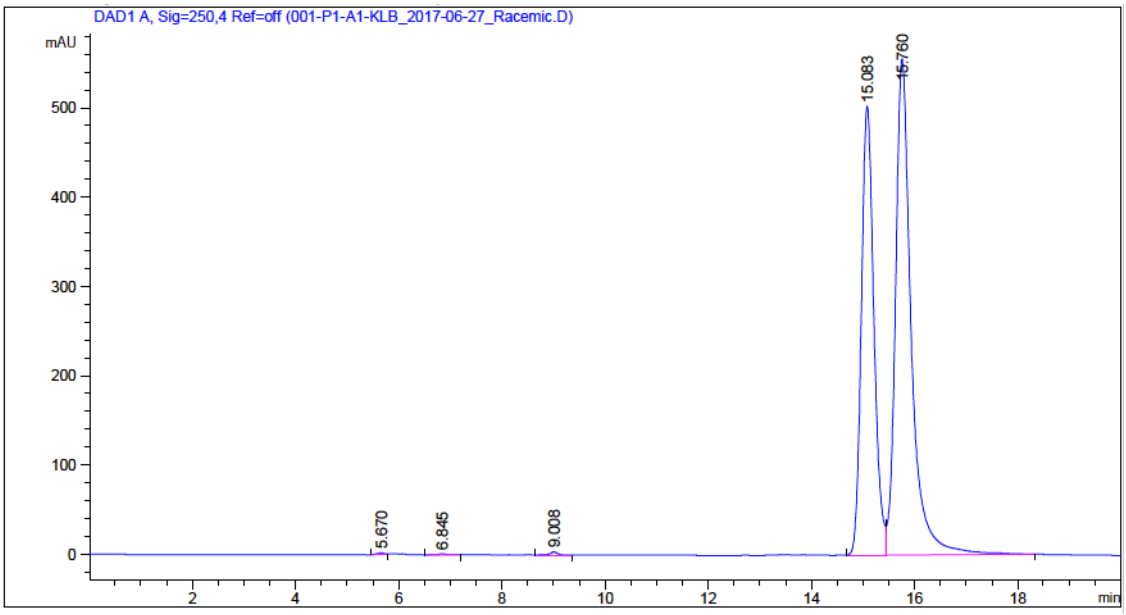


Figure 1.26: HPLC Trace of (D)-BN-Tryptophan ester **1.74**.

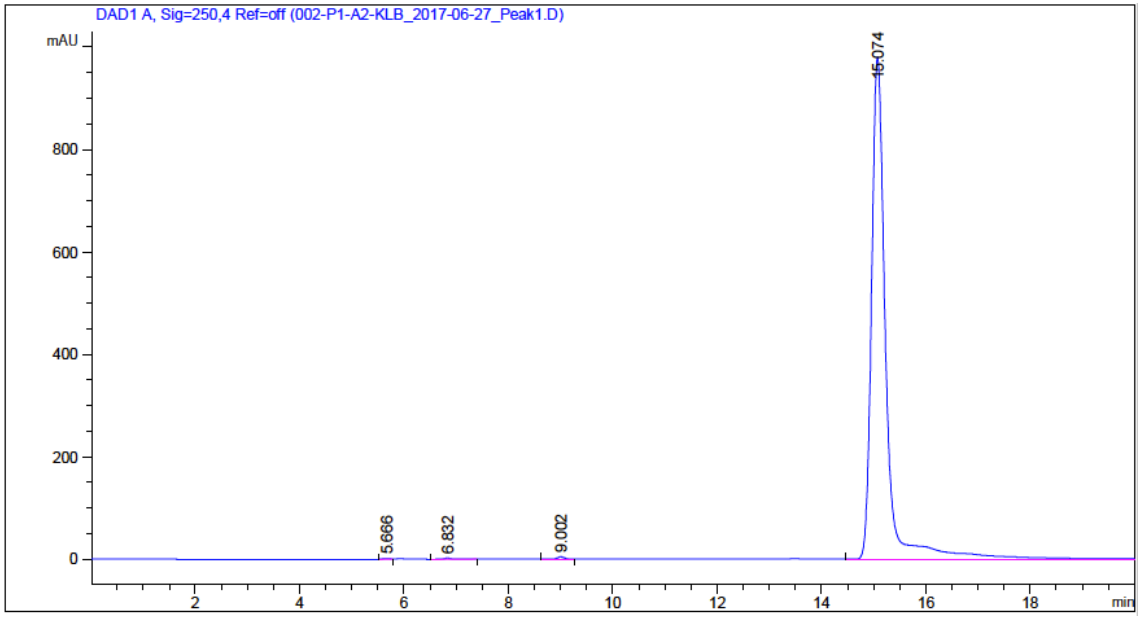
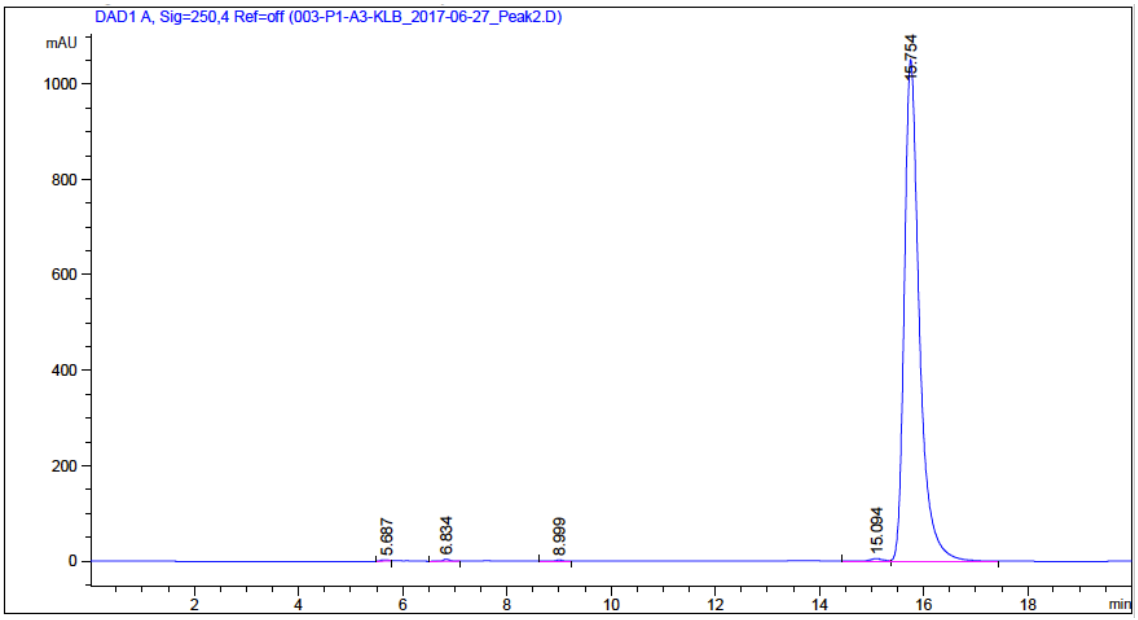
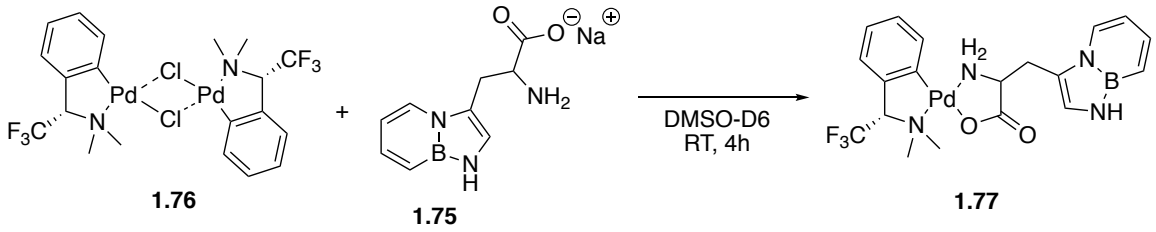


Figure 1.27: HPLC Trace of (L)-BN-tryptophan ester **1.74**.



1.5.4. Determination of Enantiopurity and Absolute Configuration of BN-tryptophan

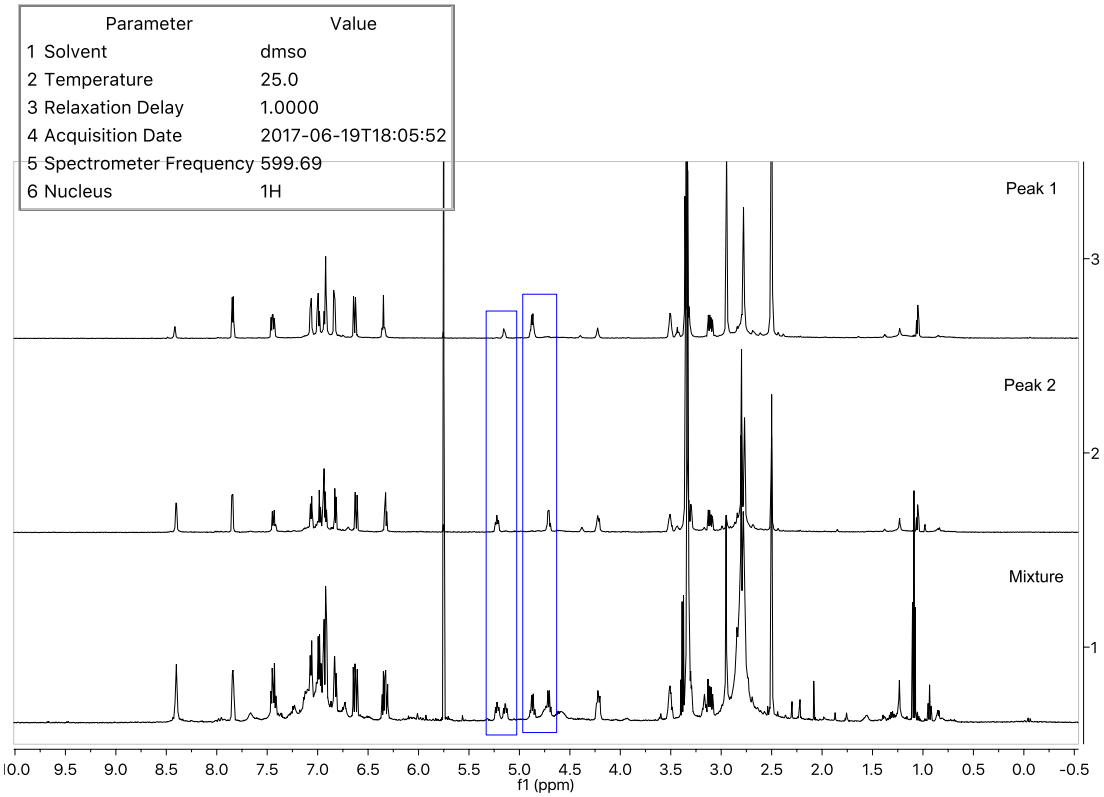
NMR Chiral Analysis:

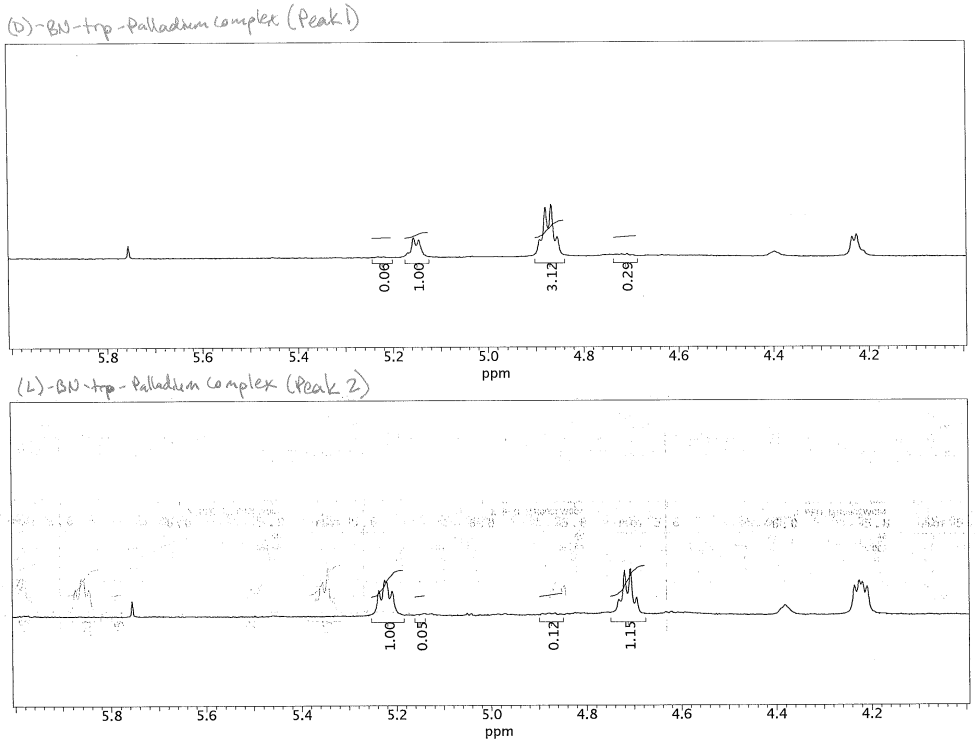


1.76 **1.75** **1.77**

A 4mL vial was charged with Pd-complex[66] **1.76** (7.7 mg, 0.011 mmol), BN-tryptophan **1.75** (5.0 mg, 0.022 mmol), and a stir bar. DMSO-D6 (0.5 mL) was added and the solution was allowed to stir under nitrogen for 4 hours, after which the NMR spectra were collected. The NMR showed quantitative formation of **1.77**. The spectra of the resolved compounds, **(D)-1.75** and **(L)-1.75** are as shown in Figure **1.28**.

Figure 1.28: NMR Analysis of Enantiopurity of Resolved BN-trp **1.75**.





e.e. calculation*:*

d.e. Peak 1 complex (= e.e. (D)-BN-trp **1.75**) = (1.00-0.06)/(1.00+0.06)= 0.8868x100% = 90%

d.e. Peak 2 complex (= e.e. (L)-BN-trp **1.75**) = (1.00-0.05)/(1.00+0.05)= 0.9048x100% = 90%

Specific Rotation:

(D)-BN-tryptophan 1.75 $[\alpha]^{25}_{D} = -7.600$ (c = 0.50, H_2O, *l* = 0.5 dm)

(L)-BN-tryptophan 1.75 $[\alpha]^{25}_{D} = +7.600$ (c = 0.50, H_2O, *l* = 0.5 dm)

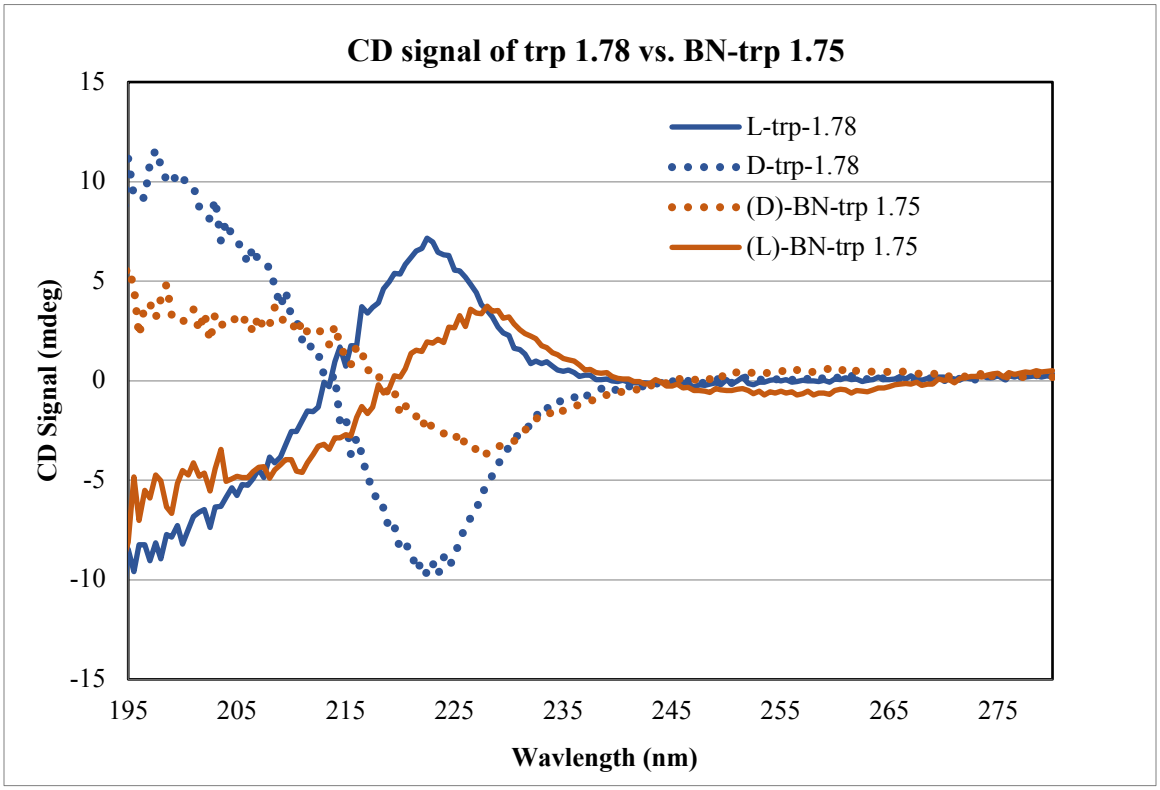
(D)-tryptophan 1.78 $[\alpha]^{25}_{D} = -3.600$ (c = 0.50, H_2O, *l* = 0.5 dm)

(L)-tryptophan 1.78 $[\alpha]^{25}_{D} = +3.200$ (c = 0.50, H_2O, *l* = 0.5 dm)

CD-spectroscopy:

A 0.1 mm quartz cuvette was used for spectroscopy measurements. Methanol was used as solvent and concentrations of 0.1 mg/mL were used. Step size was 0.5 nm and the averaging time was 6 seconds. Pure methanol was used as the blank and the CD spectra were corrected according to it (baseline subtraction). Spectra are an average of 3 scans.

Figure 1.29: CD-signal of Tryptophan **1.78** Compared to BN-trp **1.75**



1.5.5. Quantum Yield Measurement of (L)-BN-tryptophan and (L)-tryptophan

A dilute solution of each (L)-BN-Trp **1.75** and (L)-Trp **1.78** was prepared in water in a 1 cm quartz cuvette. Each solution was diluted to have an absorbance intensity between 0.5 and 0.7 a.u. The same solution was then used to measure the quantum yield with the "petite" integrating sphere.

For (L)-BN-Trp **1.75**:

Settings:
293:265-570 nm, excitation slit width = 3.0 nm, emission slit width: 5.0 nm

Results:
1) Abs = 0.493, $\Phi = 0.297$
2) Abs = 0.528, $\Phi = 0.288$
3) Abs = 0.643, $\Phi = 0.307$
4) Abs = 0.605, $\Phi = 0.286$
5) Abs = 0.558, $\Phi = 0.275$
Average: $\Phi = 0.291 \pm 0.011$

For (L)-Trp **1.78**:

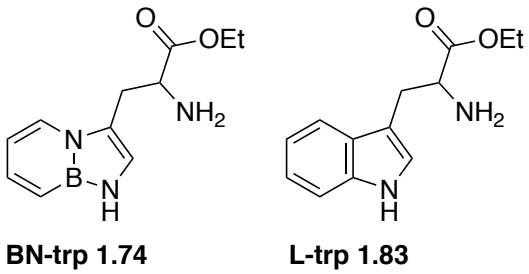
Settings:
279:250-550 nm, excitation slit width = 3.0 nm, emission slit width: 5.0 nm

Results:
1) Abs = 0.567, $\Phi = 0.251$
2) Abs = 0.628, $\Phi = 0.248$
3) Abs = 0.577, $\Phi = 0.247$
Average: $\Phi = 0.249 \pm 0.0017$

1.5.6. Solvatochromism of BN-tryptophan Ester and L-tryptophan Ester

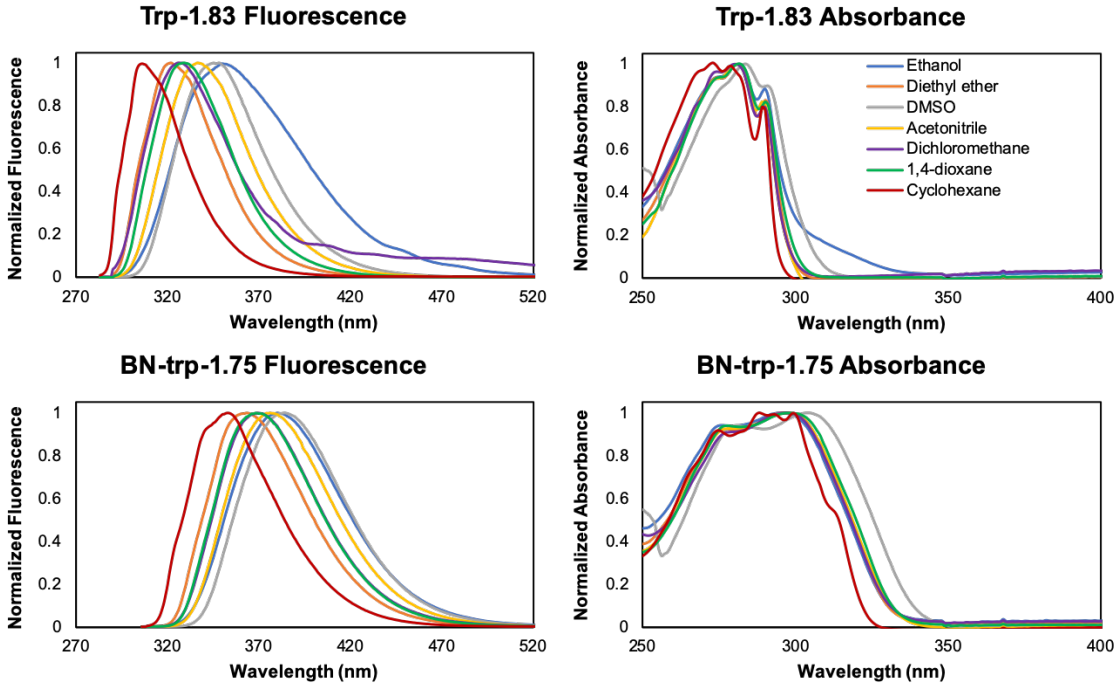
All solvents were degassed through at least 3 freeze-pump-thaw cycles and solutions were prepared in a nitrogen filled dry box.

A dilute solution (~1-1.4x10^{-4} M) of each BN-Trp ester **1.74** or (L)-Trp ester **1.83** was prepared in the specified solvent in a 1 cm quartz cuvettes and the absorbance and fluorescence were measured.



BN-trp 1.74 **L-trp 1.83**

Figure 1.30: Absorbance and Emission Spectra of Trp-**1.69** and BN-trp **1.60**



1.5.7. NMR Assignment of BN-tryptophan

Figure 1.31: gCOSY spectrum of BN-trp **1.75**

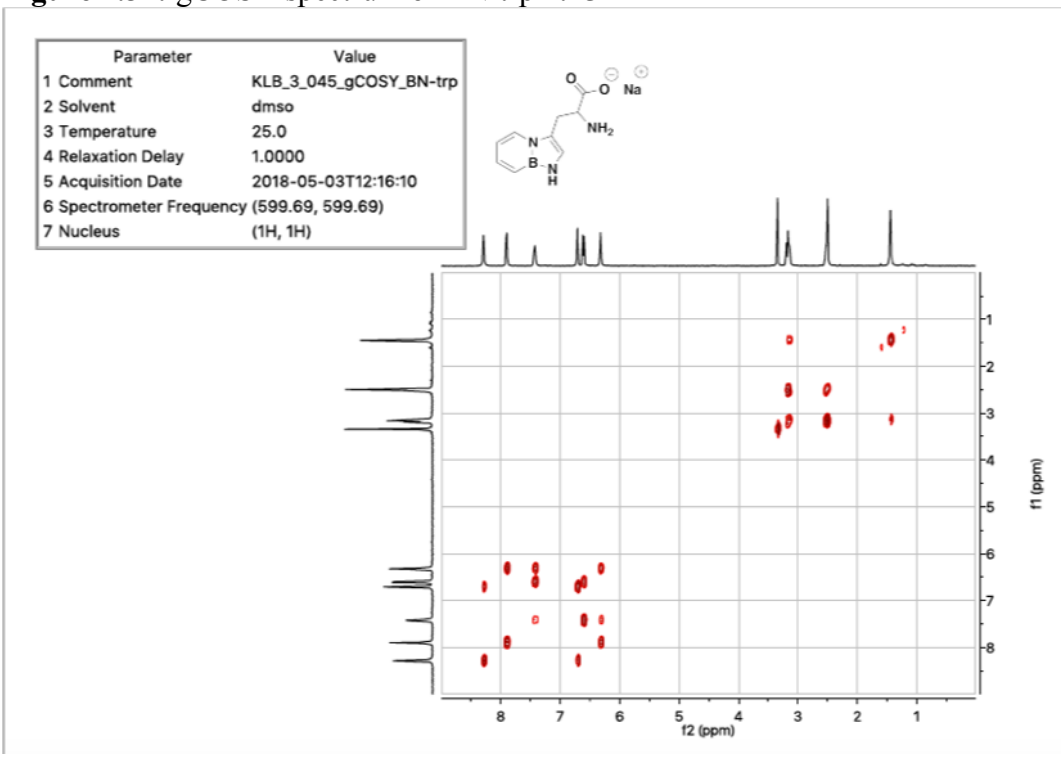


Figure 1.32: gHSQC spectrum of BN-trp **1.75**

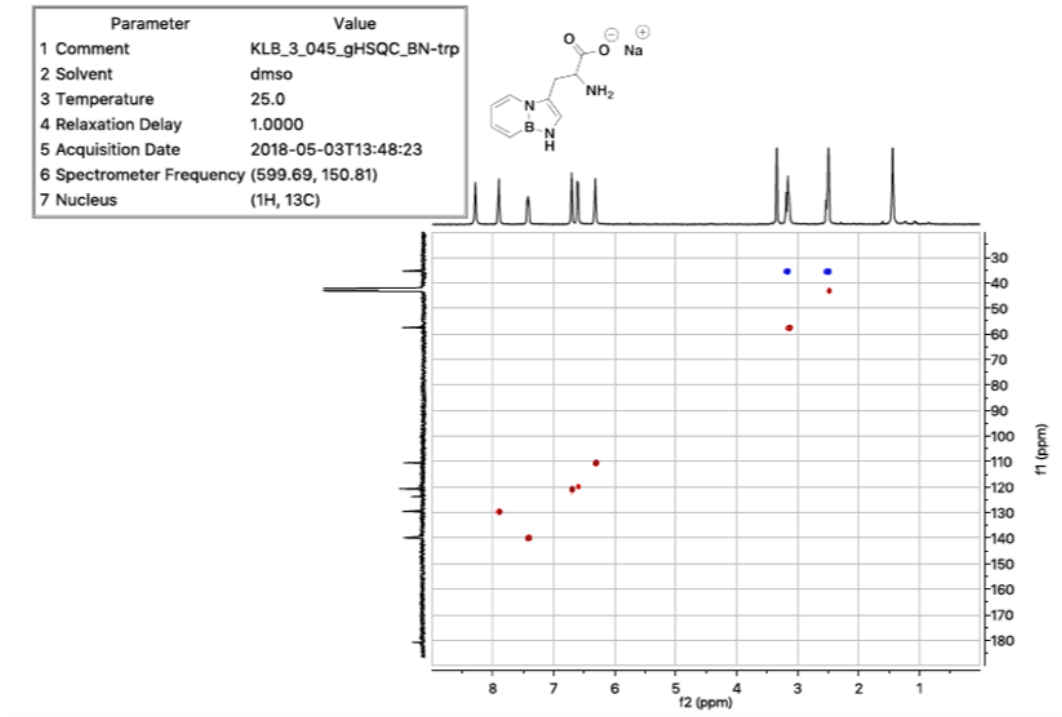


Figure 1.33: ^{1}H NMR spectrum of BN-trp **1.75** with proton assignments

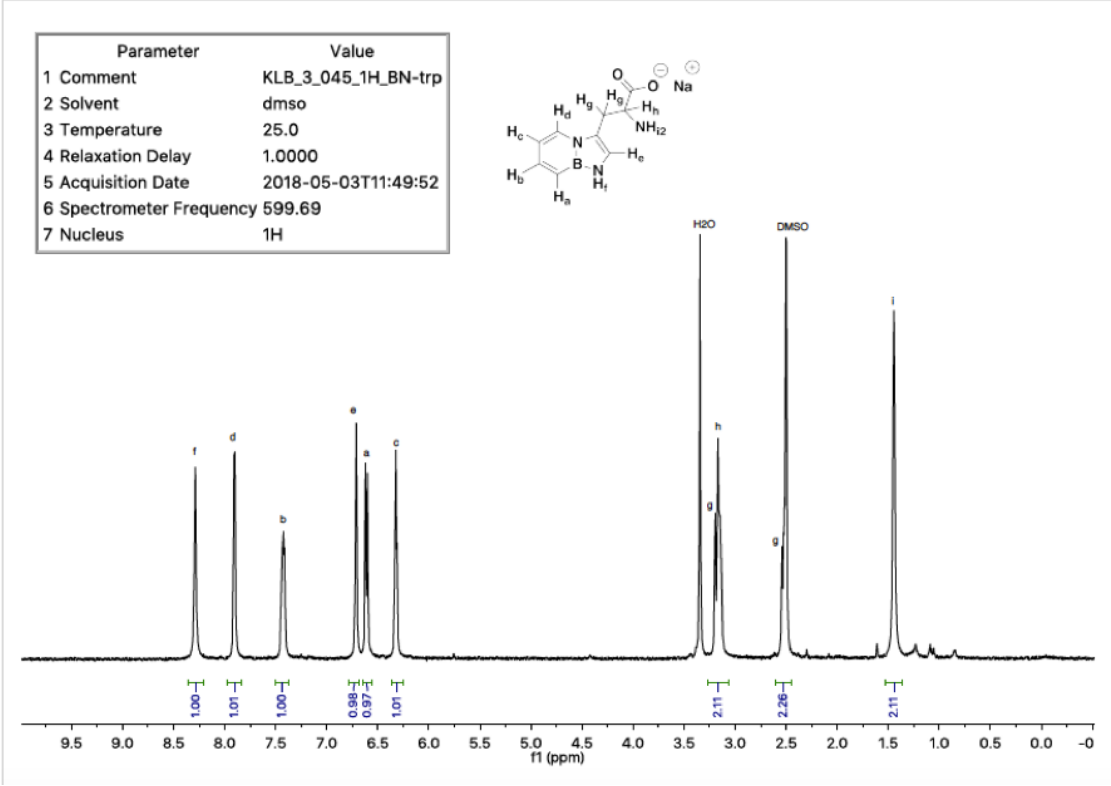
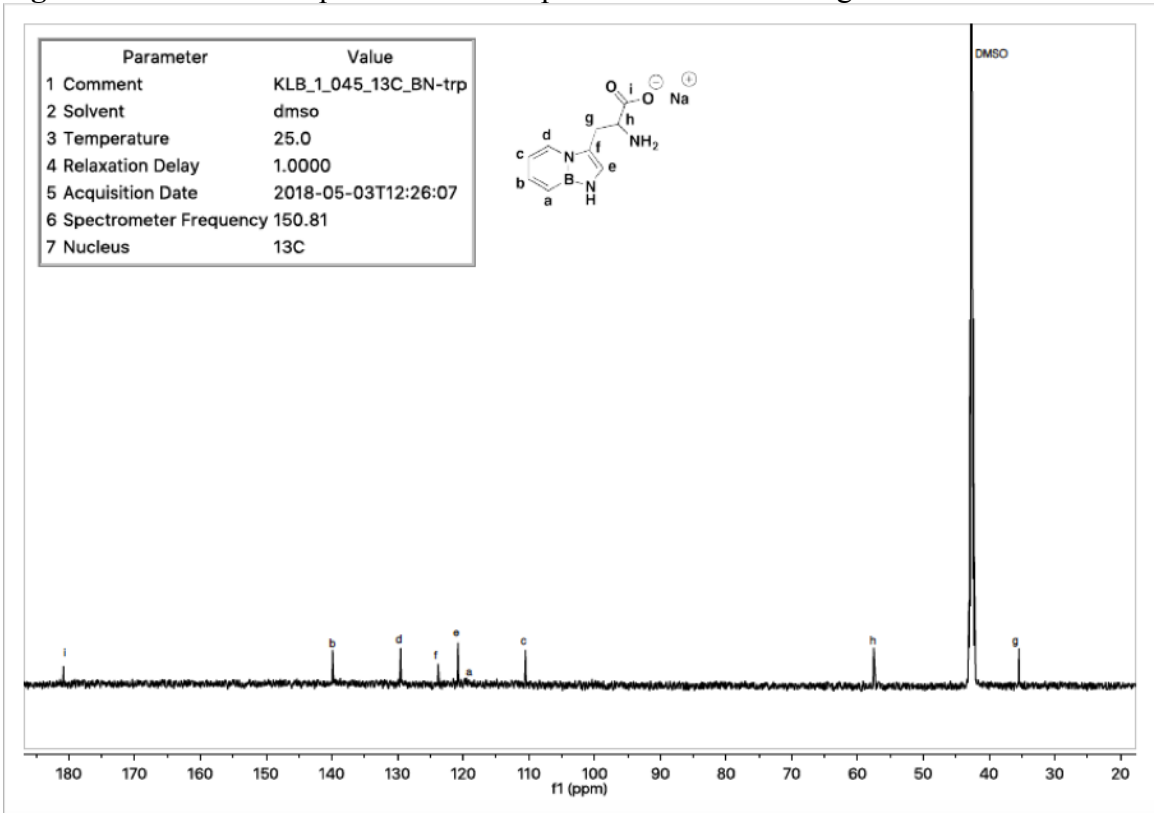


Figure 1.34: ^{13}C NMR spectrum of BN-trp **1.75** with carbon assignments



1.5.8. Cell Fermentation and Protein Isolation

Trp-auxotrophic *Escherichia coli* strain was purchased from ATCC (catalog number 49980 [genotype *WP2 uvrA*]). All fermentation and expression experiments were performed in GMML supplemented with 200 μM glucose (minimal media) (see recipe below). The expression host *E. coli* ATCC49980 was routinely transformed with either plasmid: pET22b-sfGFPwt, pET22b-sfGFP151TGG (subcloned via directed mutagenesis from pET22b-sfGFPwt using standard cloning methods, sequence below), or pET22b-Llac-KSI-7TGA.

The incorporation experiments were performed using cultures grown in minimal media in the presence of 0.0075 mM tryptophan (the optimal limiting concentration of the native substrate at 30 °C) and 2 mM Ampicillin (in the transformed cell cultures). Cell cultures were grown to ~0.5 OD_{600}, at which time tryptophan was depleted and the desired tryptophan analogue (1 or 2 mM) was added along with 1 mM IPTG. End-point optical densities and fluorescence were measured after incubation at 30 °C overnight.

Overnight expression cultures were purified as described in Italia et al.[70]

GMML Recipe:

Minimal Media Recipe

Glycerol based minimal media supplemented with leucine (GMML)

For every liter culture, make:

- 12.8 g $Na_2HPO_4 \cdot 7H_2O$ (or 6 g Na_2HPO_4)
- 3 g KH_2PO_4
- 0.5 g NaCl
- 1 g NH_4Cl

Bring to approximately 450 ml with **Tap H_2O**, and adjust pH to 7.4, and then add **Tap H_2O** to 0.5L
Autoclave the above solution
Also autoclave 10 ml glycerol in 500 ml **Tap H_2O**.
If you need to make agar plates, add 15 g agar into the above 500 mL glycerol solution before autoclaving.

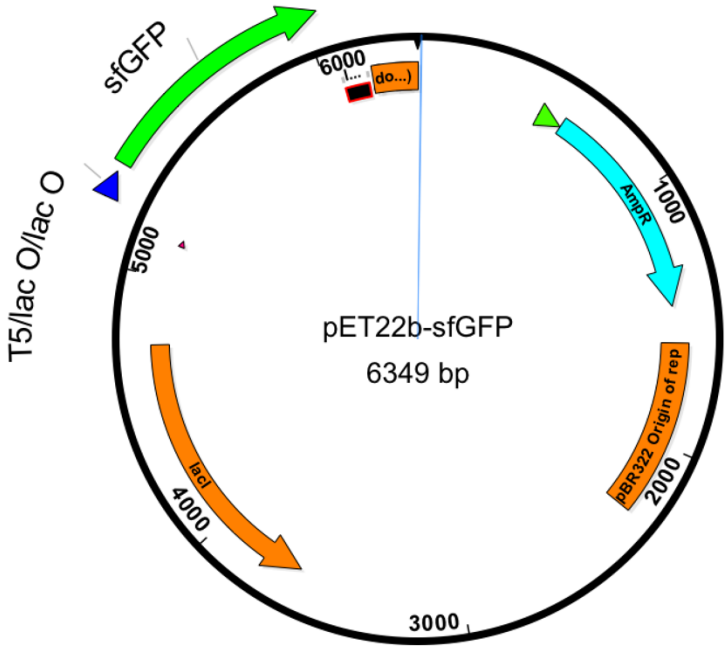
GMML media are made by mixing the above two 500 ml solutions and the following ingredient.

Make other stock solutions (~10 mL each), sterile them separately through 0.22 uM filters:

1. 300 mM L-leucine (1000x stock; adjust with minimal NaOH for solubilization)
2. 4 mM d-biotin (1000X stock)
3. 330 uM Thiamine (1000X stock)
4. 1 M $MgSO_4$ (1000X stock)
5. 0.1 M $CaCl_2$ (1000X stock)
6. Misc. heavy metal solution (1000X stock): To make 10 mL stock, add 10 mg of $CuSO_4 \cdot 5H_2O$, 40 mg $MnCl_2 \cdot 4H_2O$, 40 mg $ZnCl_2$, 12 mg $FeSO_4 \cdot 5H_2O$; store at -20C for stability

Above 6 components have to be added before use (cannot be added when the autoclaved solutions are hot).

Plasmid Sequences:



tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacactt
gccagcgccctagcgcccgctccttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctc
taaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatgg
ttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactc
ttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcct
attggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtgg
cacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaga
caataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattc
ccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagtt
gggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttt
ccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtc
gccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagt
aagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccg
aaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaag
ccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaact
acttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcc
cttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggc
cagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagat
cgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgattta
aaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagt
tttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctg
ctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaag
gtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaact
ctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttac
cgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaa
aggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta
tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagccta
tggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgt
tatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcg
cagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacac
cgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgt
gactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccg
cttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggca
gctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagt

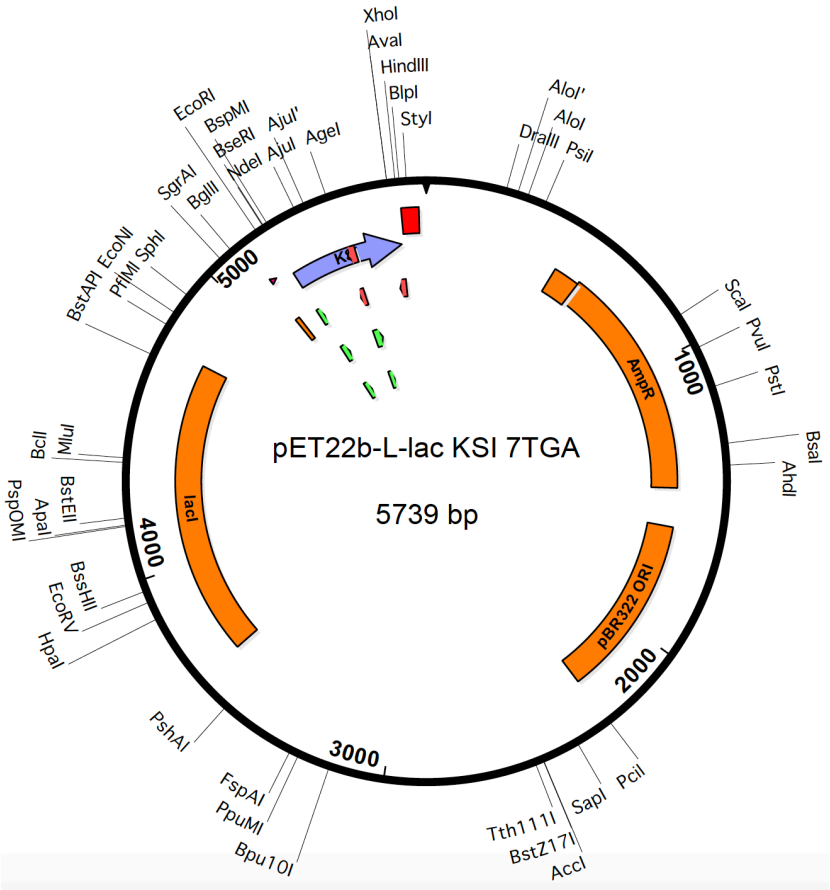
ttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatg
cctccgtgtaaggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttac
tgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaa
atcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgc
agatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattca
tgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaa
ccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgc
cggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagat
tccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgct
gccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccacc
ggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacatt
aattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcg
gggagaggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttc
accgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtgg
ttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagccc
ggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctca
ttcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgat
tgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgat
ttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatg
ggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcat
ccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgcc
gcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgac
ggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggt
tgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcac
cacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccacc
ctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcga
cgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaagga
atggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgct
catgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcg
ccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagg
ggaattgtgagcggataacaattcccctctagagtttgacagcattatcatcgatctcgagaaatcataaaaaatttat
ttgctttgtgagcggataacaattatataatagattcaattgtgagcggataacaatttcacacagaattcattaaagagg
agaaattacatATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAA
TGGGCACAAATTTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACT
ACTGGAAAACTACCTGTTCCG**TGG**CCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGG
ATCACATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGA
TGACGGGACCTACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATT
GATTTTAAAGAAGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTA**TAT**ATCACGGCAG
ACAAACAAAAGAATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCA
TTATCAACAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTT
TCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGG
ATGAGCTCTACAAAGGATCCCACCACCACCACCACCACTAAaagcttaattagctgagcttggactcctgttgatagat
ccagtaatgacctcagaactccatctggatttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatc
caagctagcttggcggcggccgcactcgagcaccaccaccaccaccactgagatccggctgctaacaaagcccgaaagg
aagctgagttggctgctgccaccgctgagcaataactagcataacccctttggggcctctaaacgggtcttgaggggttt
tttgctgaaaggaggaactatatccggat

pET22b-sfGFP-151TGG

Same plasmid backbone sequence as above, but contains sfGFP Y151W (TAT->TGG) mutation.

ATGAGCAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAAT
TTTCTGTCCGTGGAGAGGGTGAAGGTGATGCTACAAACGGAAAACTCACCCTTAAATTTATTTGCACTACTGGAAAACT
ACCTGTTCCG**TGG**CCAACACTTGTCACTACTCTGACCTATGGTGTTCAATGCTTTTCCCGTTATCCGGATCACATGAAA
CGGCATGACTTTTTCAAGAGTGCCATGCCCGAAGGTTATGTACAGGAACGCACTATATCTTTCAAAGATGACGGGACCT
ACAAGACGCGTGCTGAAGTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAGGGTATTGATTTTAAAGA
AGATGGAAACATTCTTGGACACAAACTCGAGTACAACTTTAACTCACACAATGTA**TGG**ATCACGGCAGACAAACAAAAG
AATGGAATCAAAGCTAACTTCAAAATTCGCCACAACGTTGAAGATGGTTCCGTTCAACTAGCAGACCATTATCAACAAA
ATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGTCGACACAATCTGTCCTTTCGAAAGATCC
CAACGAAAAGCGTGACCACATGGTCCTTCTTGAGTTTGTAACTGCTGCTGGGATTACACATGGCATGGATGAGCTCTAC
AAAGGATCCCACCACCACCACCACCACTAA

pET22b-Llac-KSI-7TGA




```
tggcgaatgggacgcgccctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacactt
gccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctc
taaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaacttgattagggtgatgg
ttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactc
ttgttccaaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcct
attggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaatttcaggtgg
cacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgaga
caataaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattc
ccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagtt
gggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgtttt
ccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtc
gccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagt
aagagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccg
aaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaag
ccataccaaacgacgagcgtgacaccacgatgcctgcagcaatggcaacaacgttgcgcaaactattaactggcgaact
acttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcc
cttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcagcactggggc
cagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagat
cgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgatttta
aaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagt
tttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctg
ctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaag
gtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaact
ctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttac
cgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagc
ttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaa
aggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggta
```

tctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagccta
tggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgt
tatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcg
cagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacac
cgcatatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgctacgt
gactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccg
cttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggca
gctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgcgtccagctcgttgagt
ttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggttttttcctgtttggtcactgatg
cctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggttac
tgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaa
atcactcagggtcaatgccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgc
agatccggaacataatggtgcagggcgctgacttccgcgtttccagactttacgaaacacggaaaccgaagaccattca
tgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcgtatcggtgattcattctgctaa
ccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgccatgc
cggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagat
tccgaataccgcaagcgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgct
gccggcacctgtcctacgagttgcatgataaagaagacagtcataagtgcggcgacgatagtcatgccccgcgcccacc
ggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccggtgcctaatgagtgagctaacttacatt
aattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcg
gggagaggcggtttgcgtattgggcgccagggtggttttttcttttcaccagtgagacgggcaacagctgattgcccttc
accgcctggccctgagagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtgg
ttaacggcgggatataacatgagctgtcttcggtatcgtcgtatcccactaccgagatatccgcaccaacgcgcagccc
ggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaaccagcatcgcagtgggaacgatgccctca
ttcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatttgat
tgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgat
ttgctggtgacccaatgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatg
ggtgtctggtcagagacatcaagaaataacgccggaacattagtgcaggcagcttccacagcaatggcatcctggtcat
ccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgcaccgccgctttacaggcttcgacgcc
gcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaatttgcgac
ggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgactgtttgcccgccagttgttgtgccacgcggt
tgggaatgtaattcagctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcac
cacgcgggaaacggtctgataagagacaccggcatactctgcgacatcgtataacgttactggtttcacattcaccacc
ctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattcgatggtgtccgggatctcga
cgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaagga
atggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgct
catgagcccgaagtggcgagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcg
ccggtgatgccggccacgatgcgtccggcgtagaggatcgagatctcgatcccgcgaaattaatacgactcactatagg
ggaattgtgagcggataacaattgacattgtgagcggataacaagatactgagcacagaattcattaaagaggaggaat
tacatATGAACCTGCCGACCGCCtatGAAGTCCAGGGTCTGATGGCGCGTTTTATTGAACTGGTCGATGTGGGCGACAT
TGAAGCAATTGTGCAGATGTACGCGGACGATGCGACCGTTGAAAGCCCGTTTGGTCAACCACCGATTCATGGCCGTGAA
CAGATTGCGGCGCACTACCGTCAGTGGCTGGGCGGGGGTAAACTGCGTGTTTGCCTGACCGGTCCAGTGCGTGCTAGTc
atAACGGCTGCGGCGCGATGCCGTTGCGTAAAGAGTGGGTTTGGAATGGTCAGCCTTGCGCAACGGATGTTATTCTGGT
TATGCGCTTCGATGAACACGGTCGCATCCAGACCGAACAGCGCTATTGGAGCGAAGTGAATCTGAGTGTACGCGAACCG
CAGGGCAGTCTCGAGCACCACCACCACCACCACTAAgcttaattagctgagcaataactagcataacccctttggggcct
ctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggat

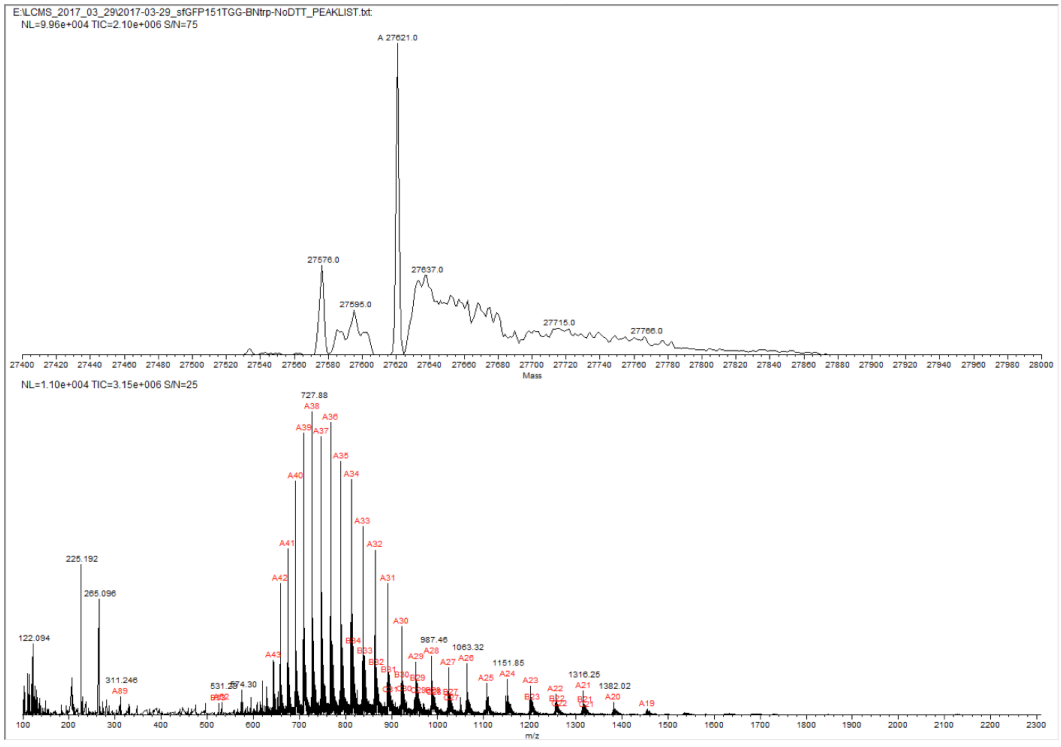
1.5.9. Oxidation Study

Protein oxidation studies were performed on the L-trp and BN-trp containing sfGFP151TGG proteins. The isolated proteins were incubated in 1 mM H_2O_2 for 5 minutes and then quenched with 1 mM DTT (dithiolthreitol). Samples were analyzed by LC/MS (Agilent Technologies, 1260 Inifinity ESI–TOF) with a water/acetonitrile gradient (95% water to 5% water) on a C18 column.

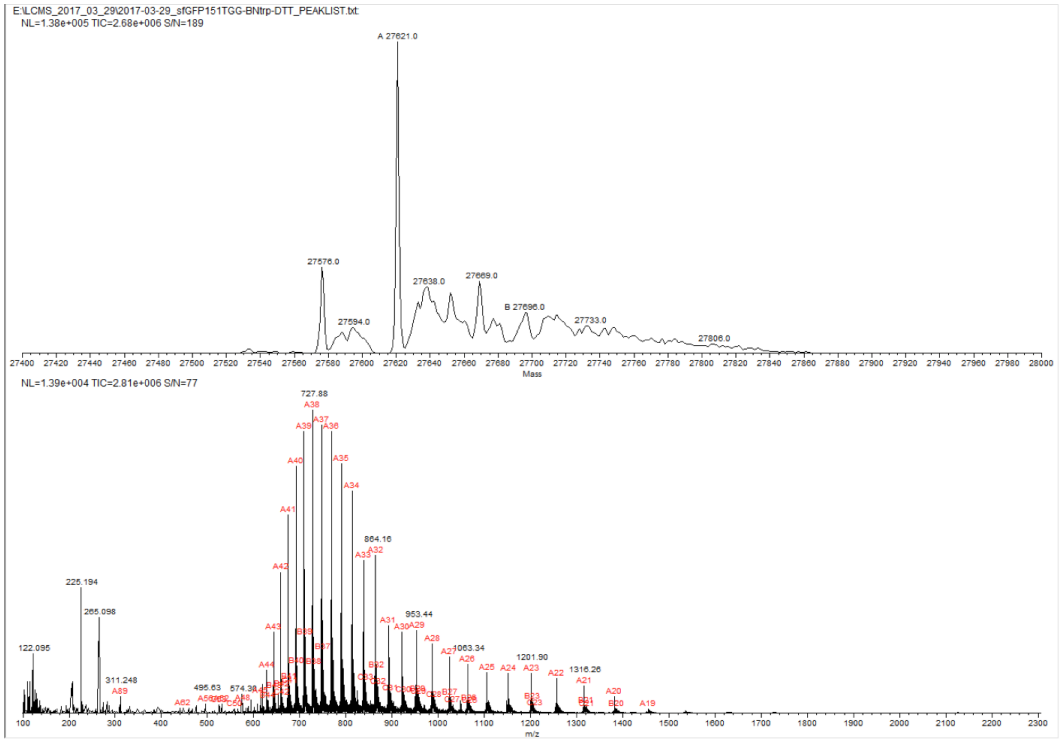
The expected mass is 27620 g/mol.

LC/MS Spectra of Oxidation Study:

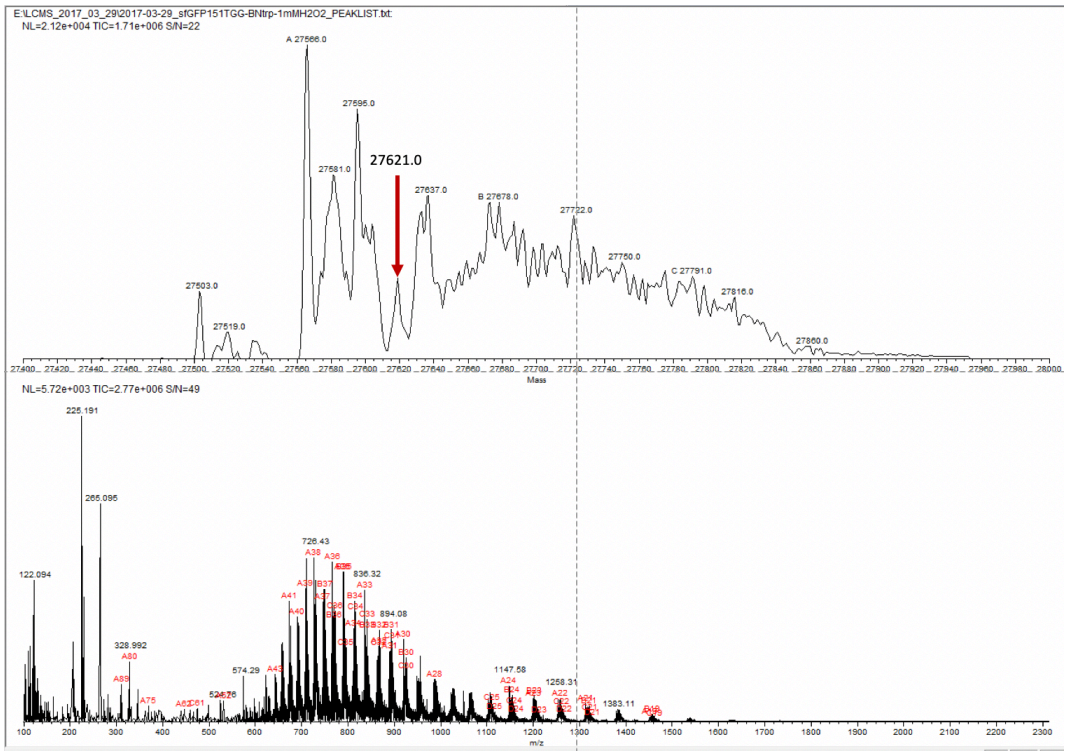
sfGFP151TGG-BN-trp: No additives



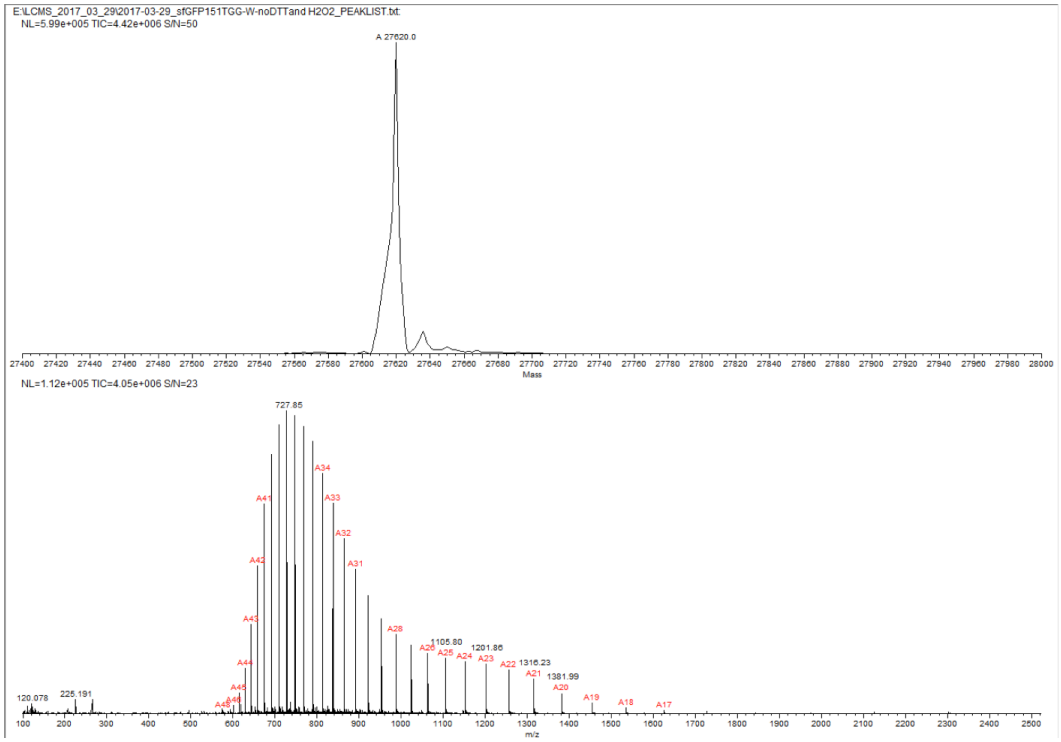
sfGFP151TGG-BN-trp: 1mM DTT



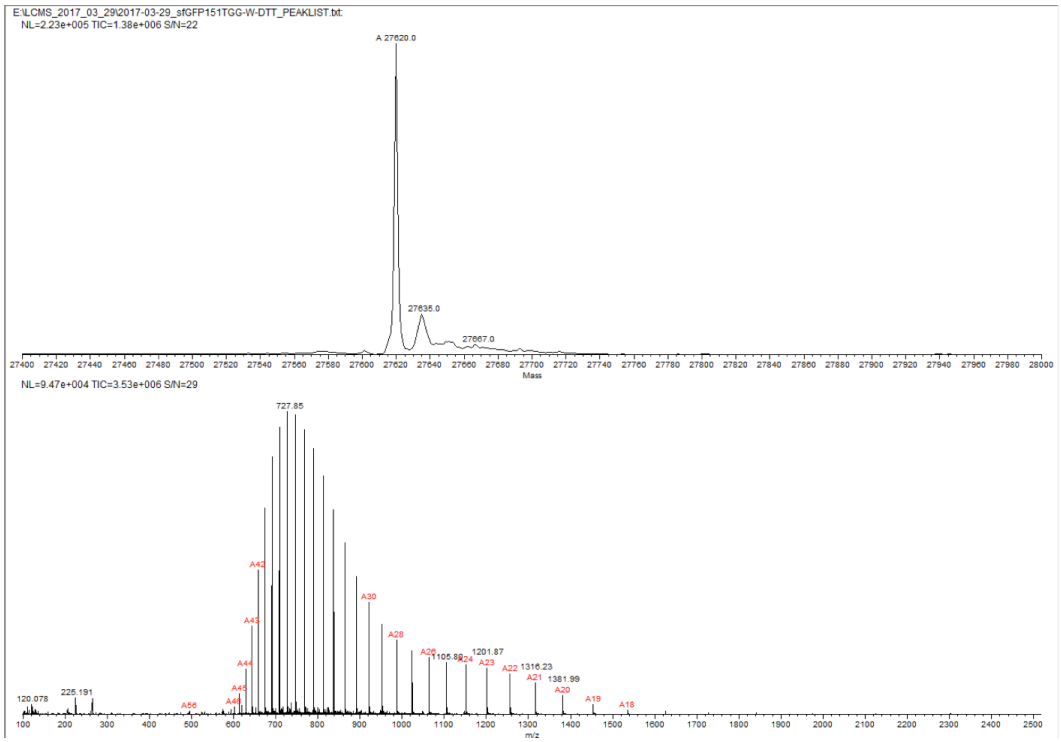
sfGFP151TGG-BN-trp: 1 mM H_2O_2 and 1mM DTT



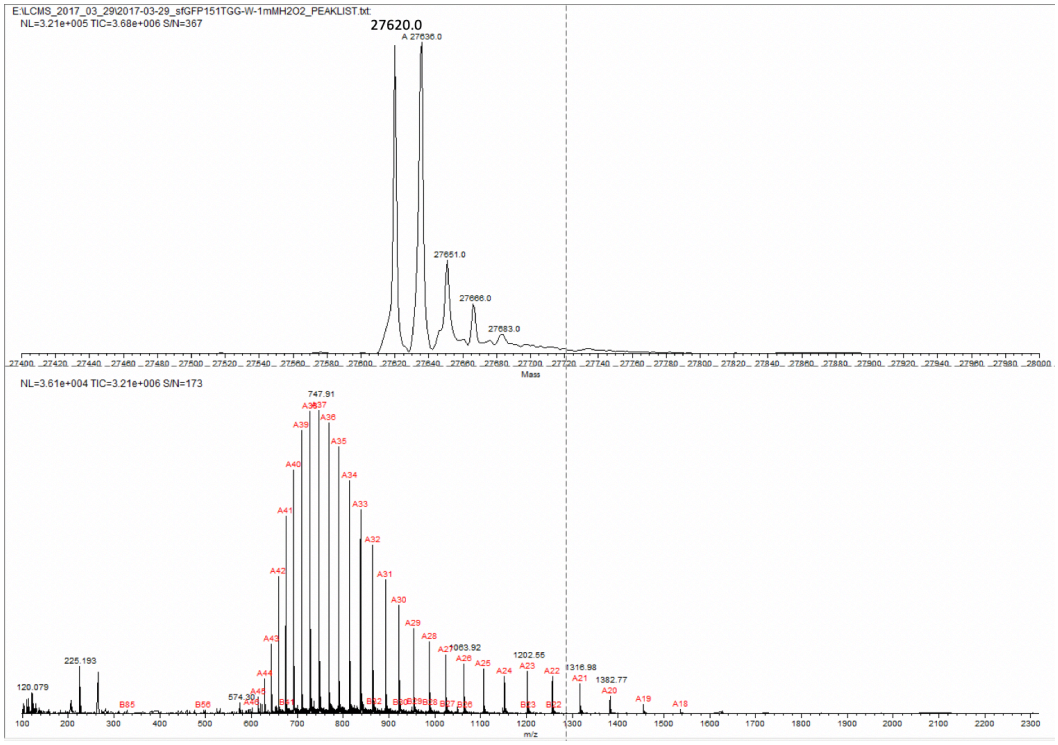
sfGFP151TGG-L-trp: No additives



sfGFP151TGG-L-trp: 1mM DTT



sfGFP151TGG-L-trp: 1 mM H_2O_2 and 1mM DTT

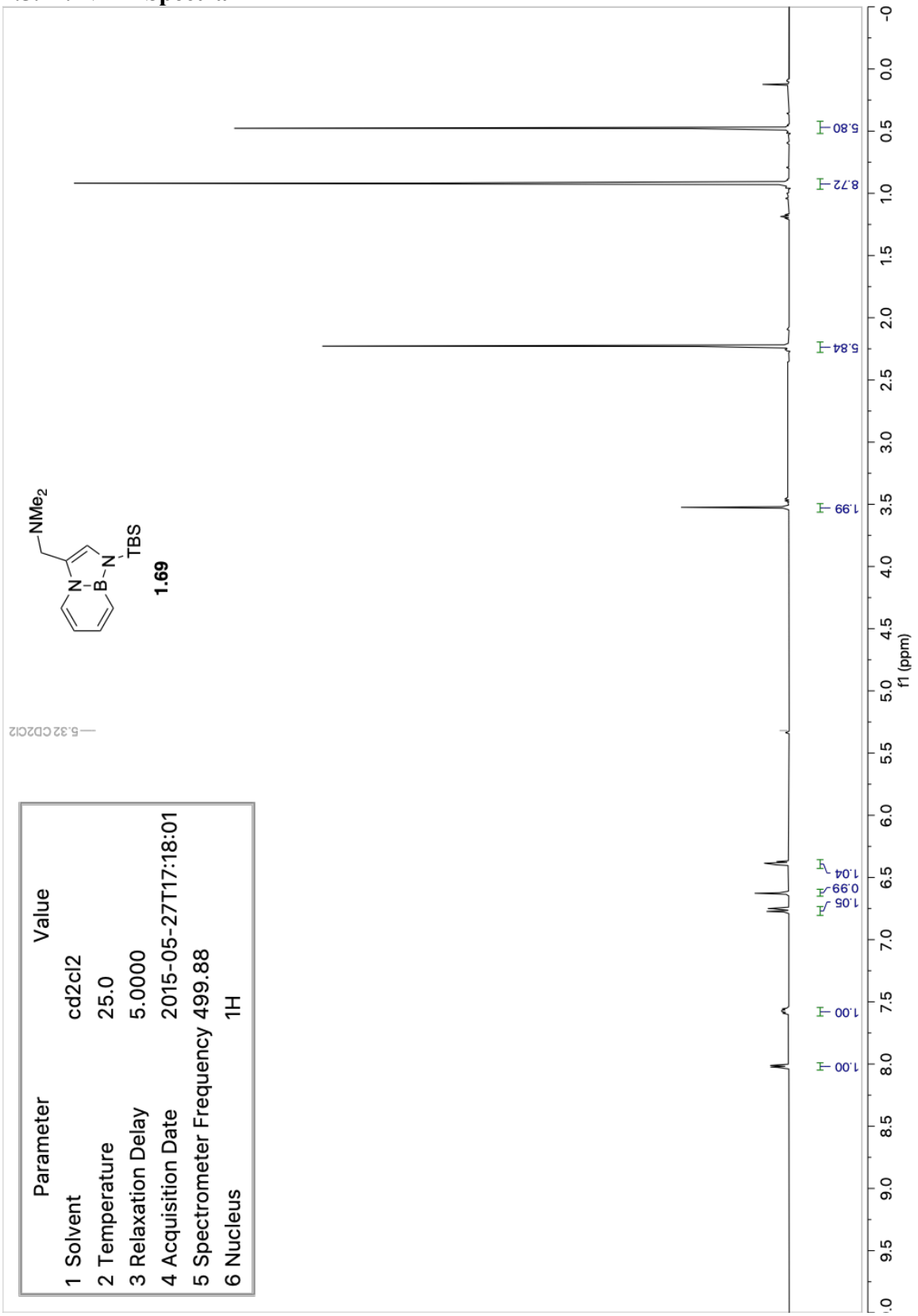


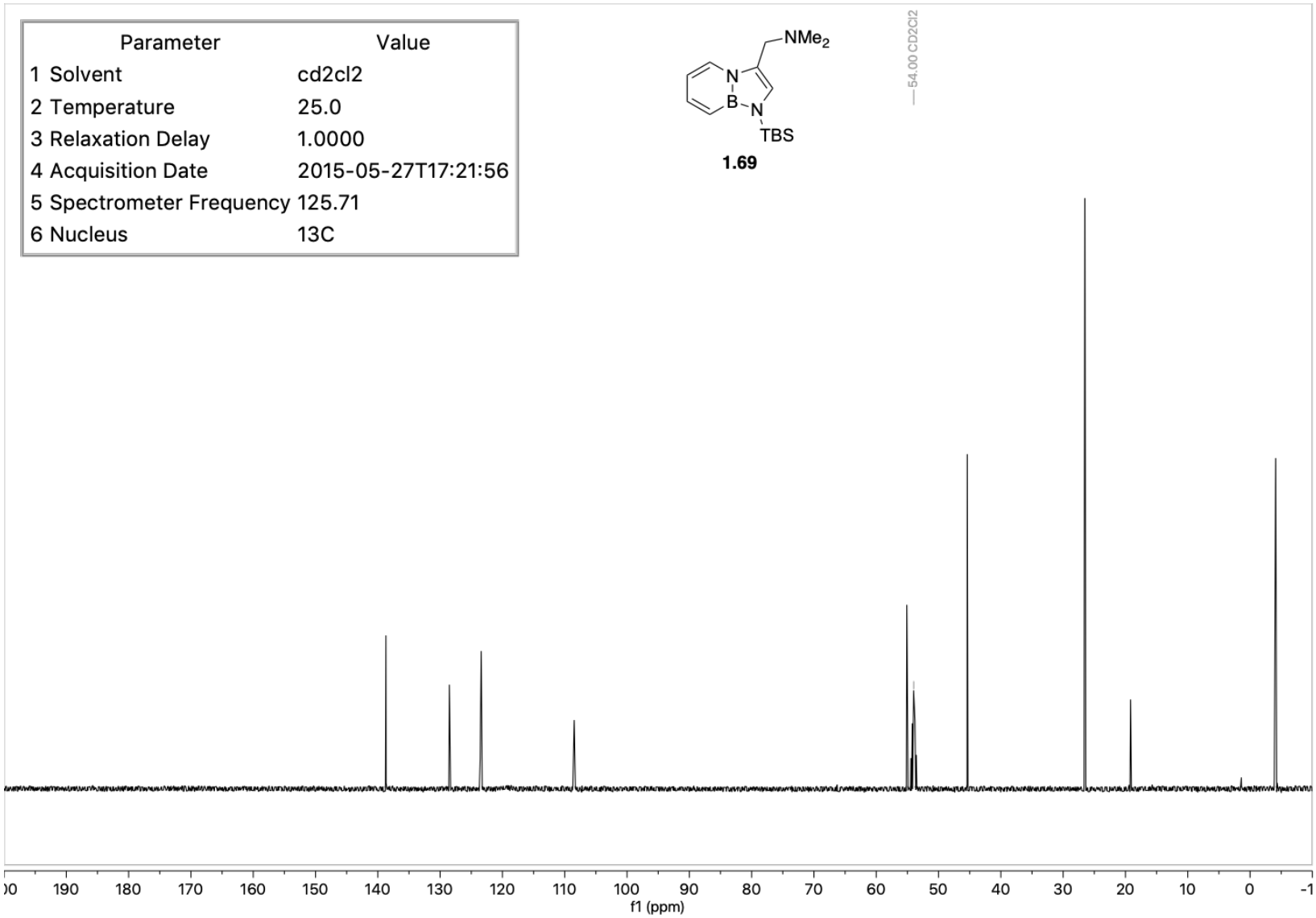
1.5.10. X-Ray Crystallographic Data for BN-tryptophan Ester

Crystal Data and Structure Refinement

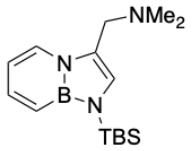
Identification code	(C17H31BN3O2Si)+(CF3O3S)-(H2O)0.5	
Empirical formula	C18 H32 B F3 N3 O5.50 S Si	
Formula weight	506.42	
Temperature	100(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	$P2_1/n$	
Unit cell dimensions	a = 15.3865(16) Å	α= 90°.
	b = 6.5658(7) Å	β= 90.703(3)°.
	c = 26.361(3) Å	γ = 90°.
Volume	2662.9(5) $Å^3$	
Z	4	
Density (calculated)	1.263 Mg/m^3	
Absorption coefficient	0.220 mm^{-1}	
F(000)	1068	
Crystal size	0.300 x 0.130 x 0.050 mm^3	
Theta range for data collection	1.524 to 28.296°.	
Index ranges	-20<=h<=20, -8<=k<=8, -35<=l<=0	
Reflections collected	12647	
Independent reflections	6615 [R(int) = 0.0426]	
Completeness to theta = 25.242°	99.9 %	
Absorption correction	Semi-empirical from equivalents	
Max. and min. transmission	0.7457 and 0.6264	
Refinement method	Full-matrix least-squares on F^2	
Data / restraints / parameters	6615 / 4 / 325	
Goodness-of-fit on F^2	1.110	
Final R indices [I>2sigma(I)]	R1 = 0.0532, wR2 = 0.1440	
R indices (all data)	R1 = 0.0836, wR2 = 0.1562	
Extinction coefficient	na	
Largest diff. peak and hole	0.575 and -0.580 $e.Å^{-3}$	

1.5.11. NMR Spectra

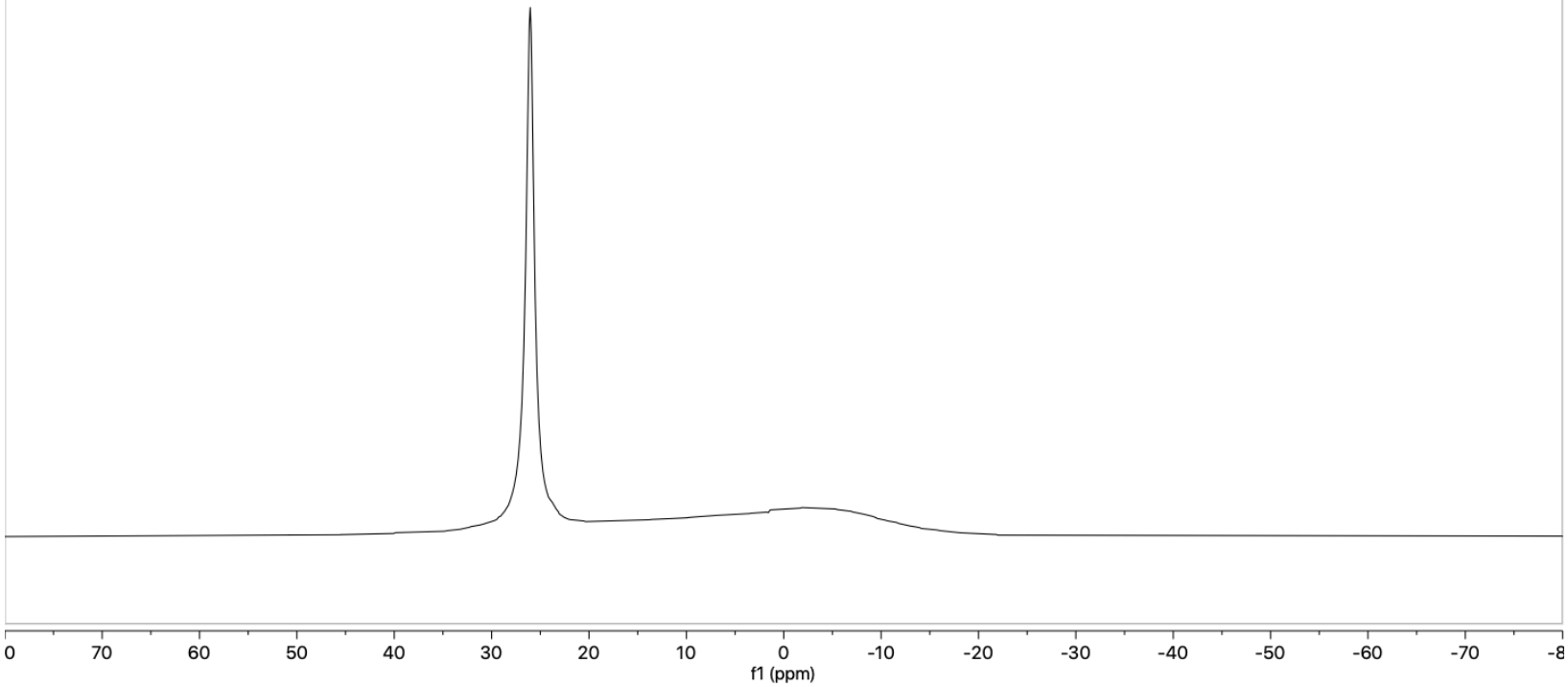



Parameter Value
1 Solvent cd2cl2
2 Temperature 25.0
3 Relaxation Delay 1.0000
4 Acquisition Date 2015-05-27T17:21:56
5 Spectrometer Frequency 125.71
6 Nucleus 13C
NMe2
N
B
N
TBS
1.69
54.00 CD2Cl2
190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -1
f1 (ppm)

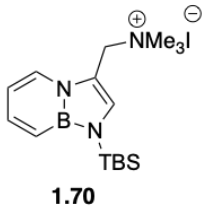
Parameter	Value
1 Solvent	cd2cl2
2 Temperature	25.0
3 Relaxation Delay	0.1000
4 Acquisition Date	2015-05-27T17:14:05
5 Spectrometer Frequency	160.38
6 Nucleus	11B



1.69

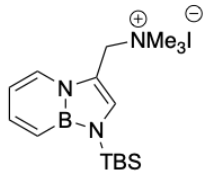


Parameter	Value
1 Solvent	dmso
2 Temperature	25.0
3 Relaxation Delay	1.0000
4 Acquisition Date	2017-06-12T15:07:30
5 Spectrometer Frequency	499.89
6 Nucleus	1H



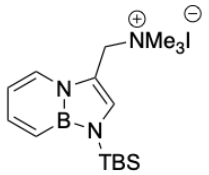
1.70

Parameter	Value
1 Solvent	dmso
2 Temperature	25.0
3 Relaxation Delay	1.0000
4 Acquisition Date	2016-10-14T12:49:58
5 Spectrometer Frequency	150.81
6 Nucleus	13C

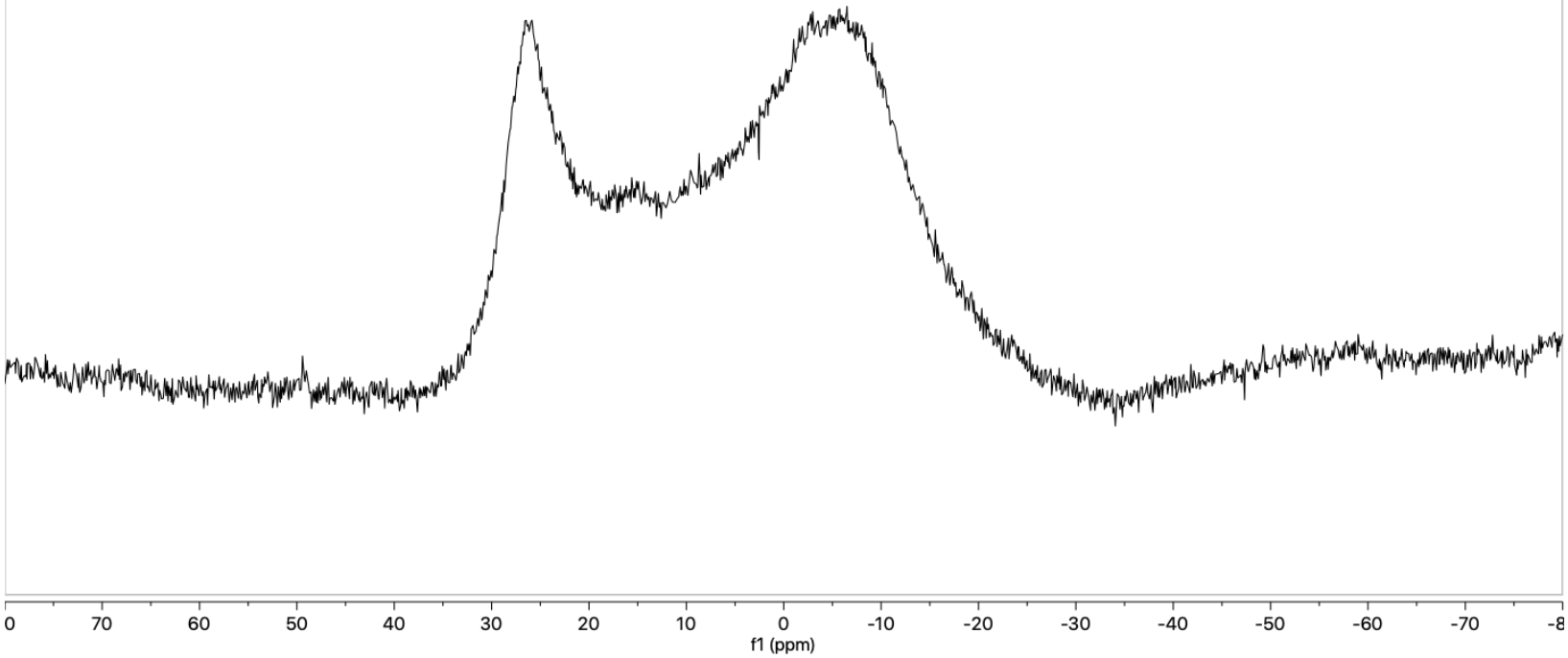


1.70

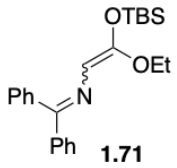
Parameter	Value
1 Solvent	dmso
2 Temperature	25.0
3 Relaxation Delay	0.1000
4 Acquisition Date	2017-06-12T15:04:40
5 Spectrometer Frequency	160.38
6 Nucleus	11B



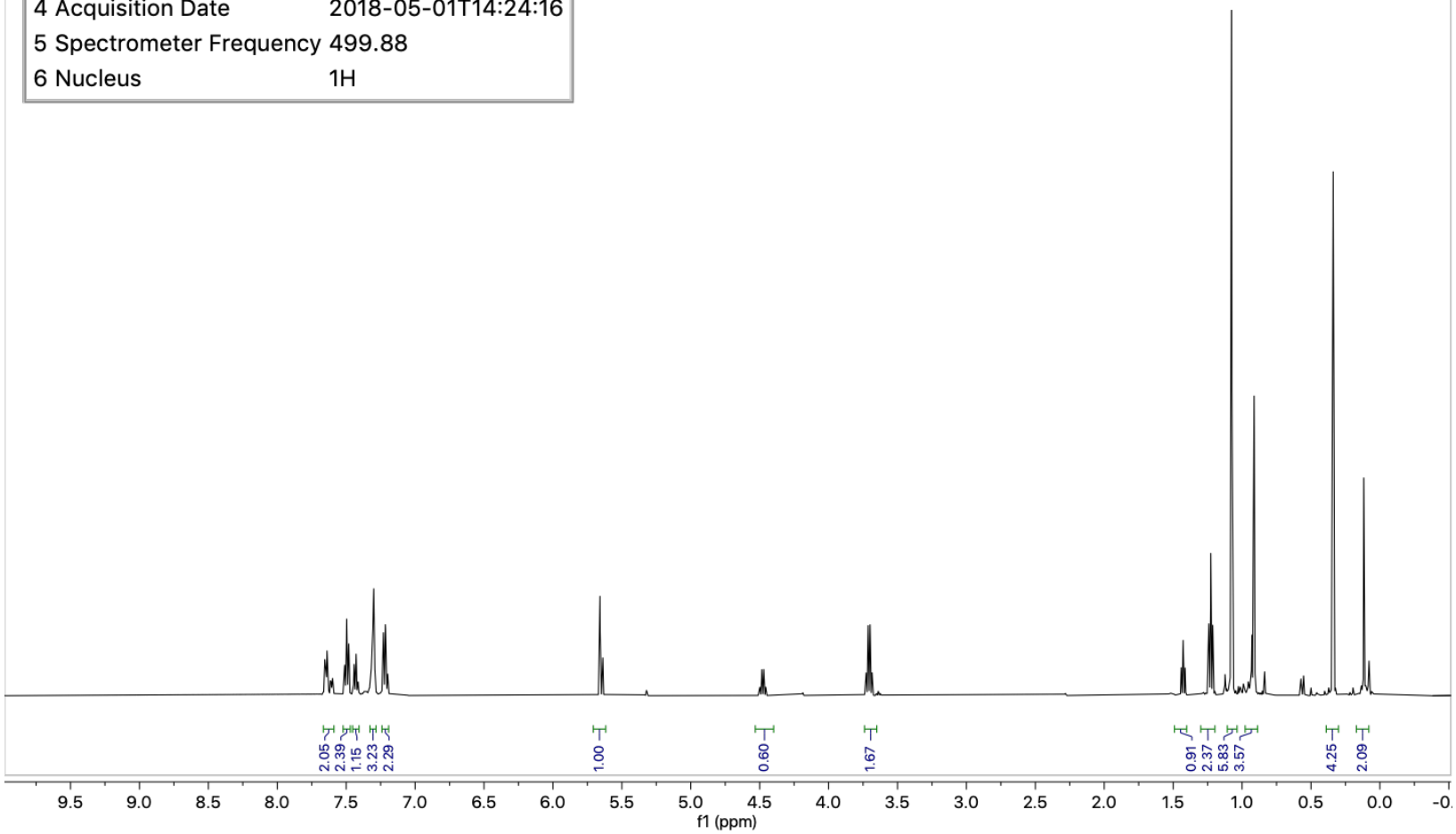
1.70



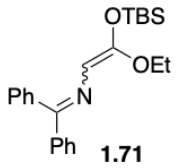
Parameter	Value
1 Solvent	cd2cl2
2 Temperature	25.0
3 Relaxation Delay	5.0000
4 Acquisition Date	2018-05-01T14:24:16
5 Spectrometer Frequency	499.88
6 Nucleus	1H



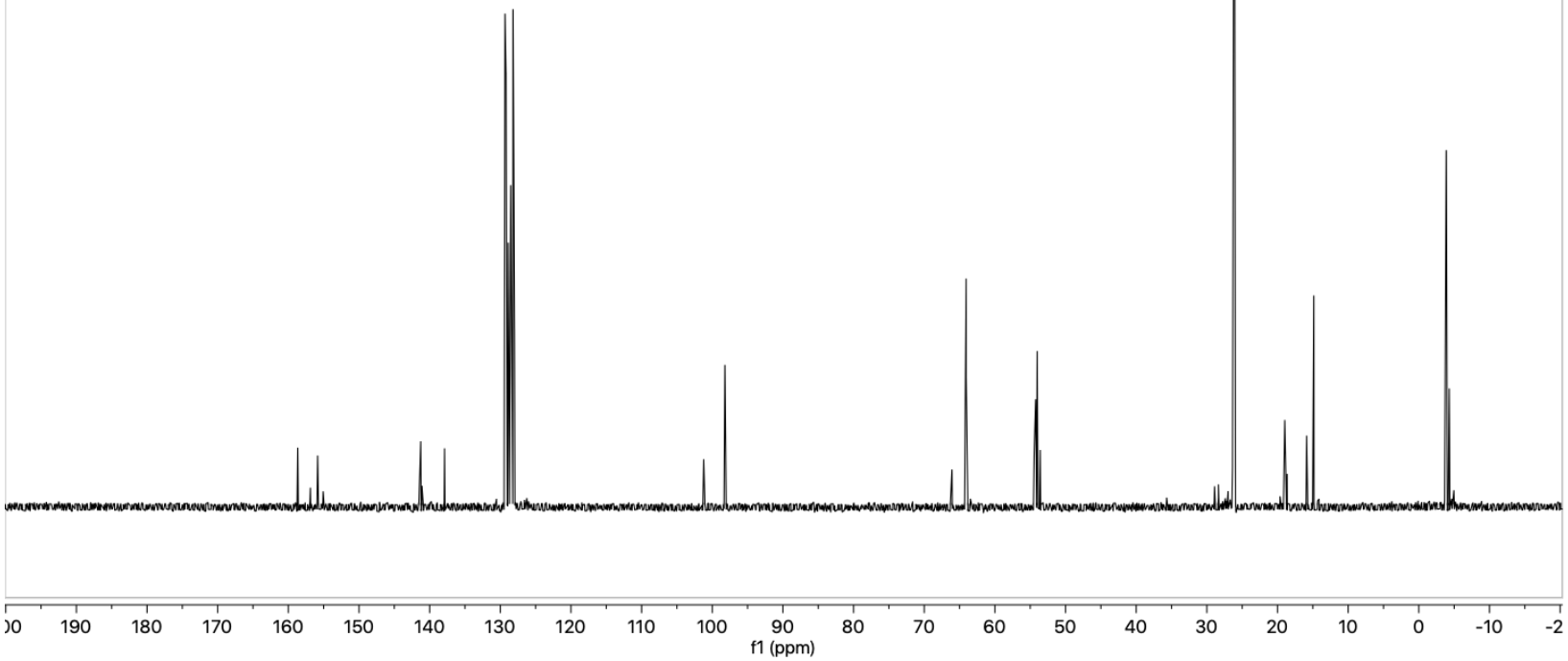
1.71

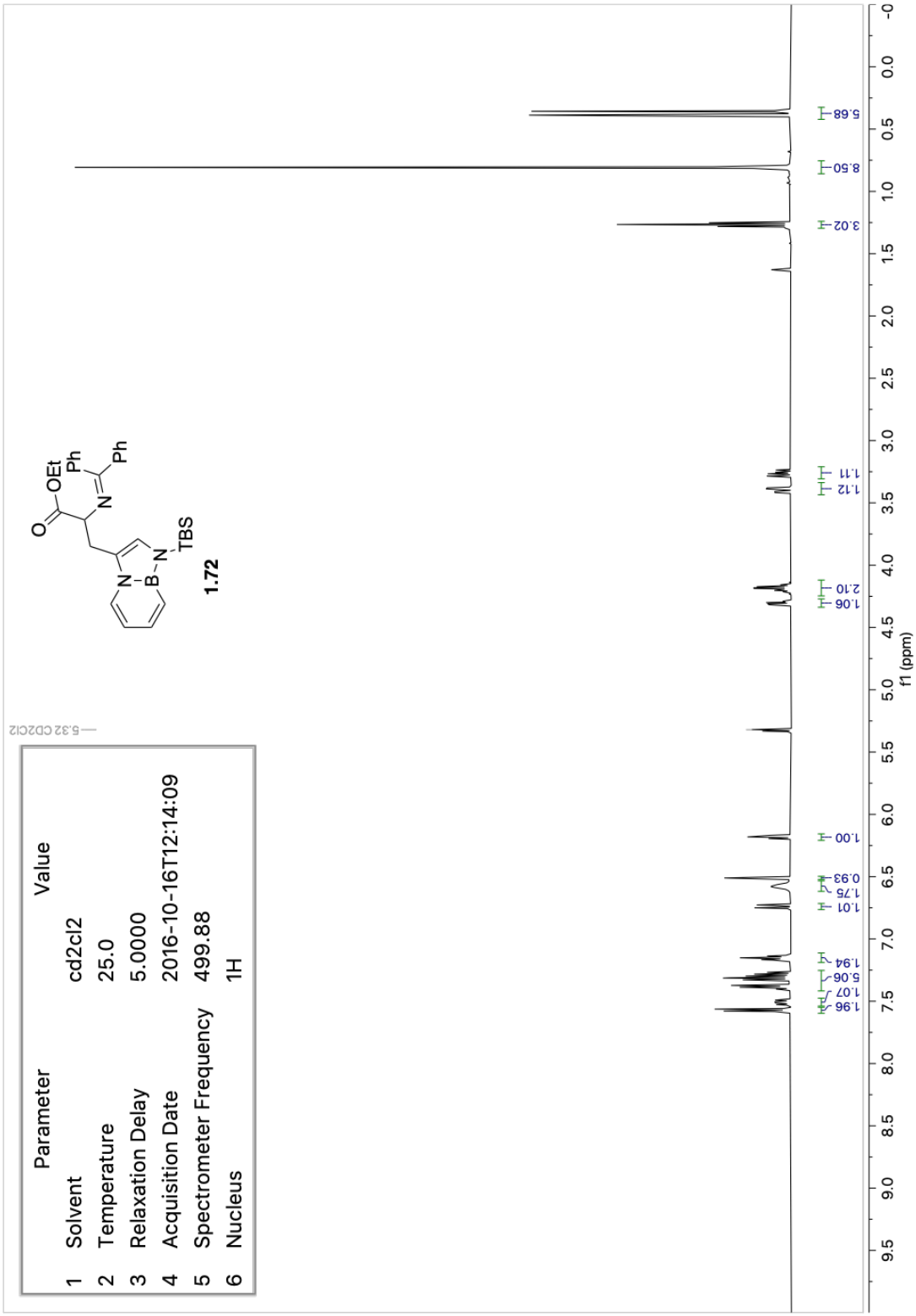


Parameter	Value
1 Solvent	cd2cl2
2 Temperature	25.0
3 Relaxation Delay	1.0000
4 Acquisition Date	2018-05-01T14:29:05
5 Spectrometer Frequency	125.71
6 Nucleus	13C

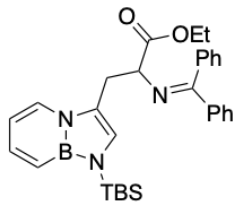


1.71



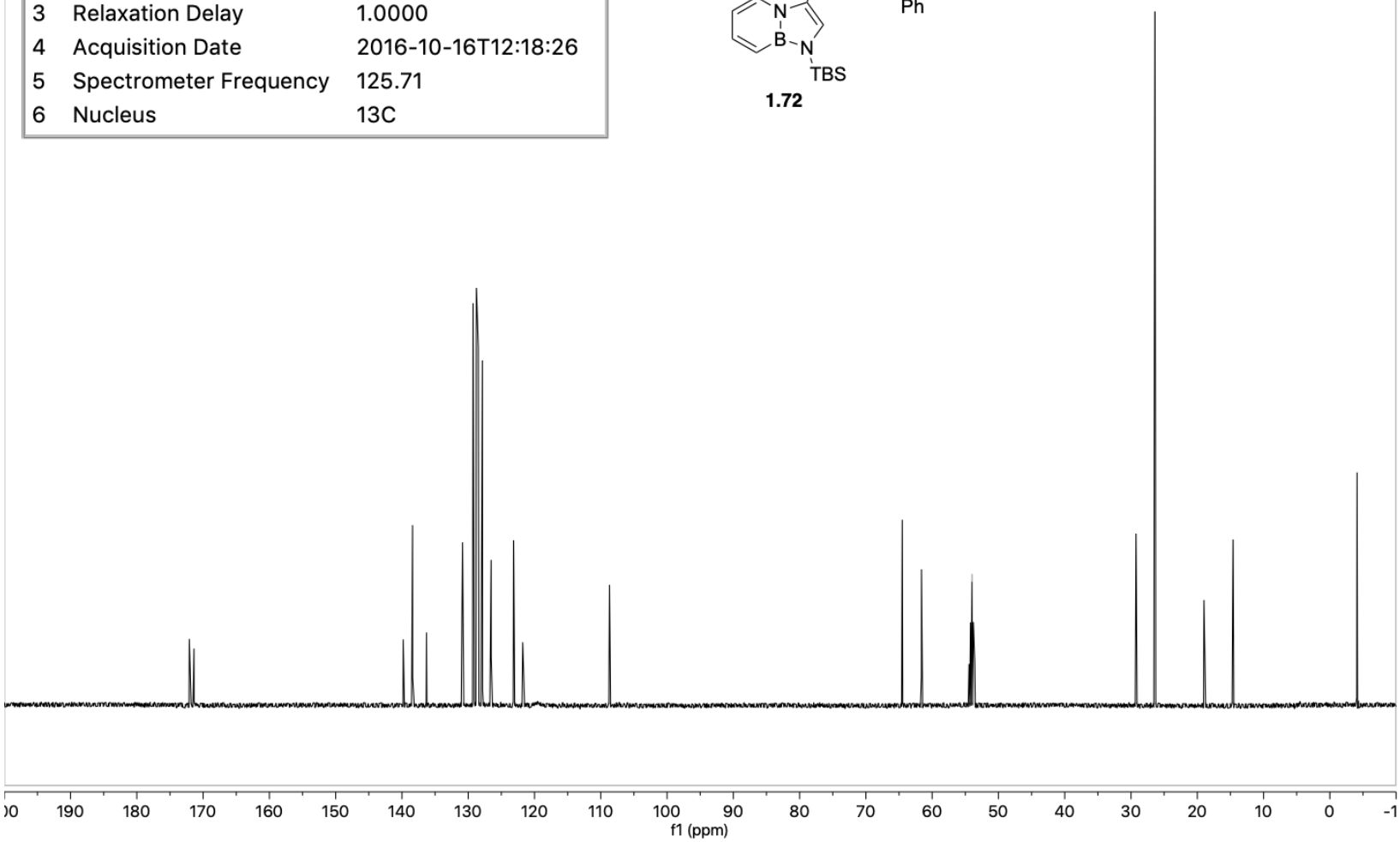

Parameter Value
1 Solvent cd2cl2
2 Temperature 25.0
3 Relaxation Delay 5.0000
4 Acquisition Date 2016-10-16T12:14:09
5 Spectrometer Frequency 499.88
6 Nucleus 1H
5.32 CD2Cl2
OEt
Ph
Ph
N
O
N
B
N
TBS
1.72
5.68
8.50
3.02
1.12
1.11
2.10
1.06
1.00
0.93
1.75
1.01
1.94
5.06
1.07
1.96
f1 (ppm)

	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	1.0000
4	Acquisition Date	2016-10-16T12:18:26
5	Spectrometer Frequency	125.71
6	Nucleus	13C

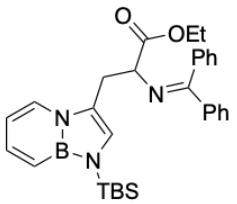


1.72

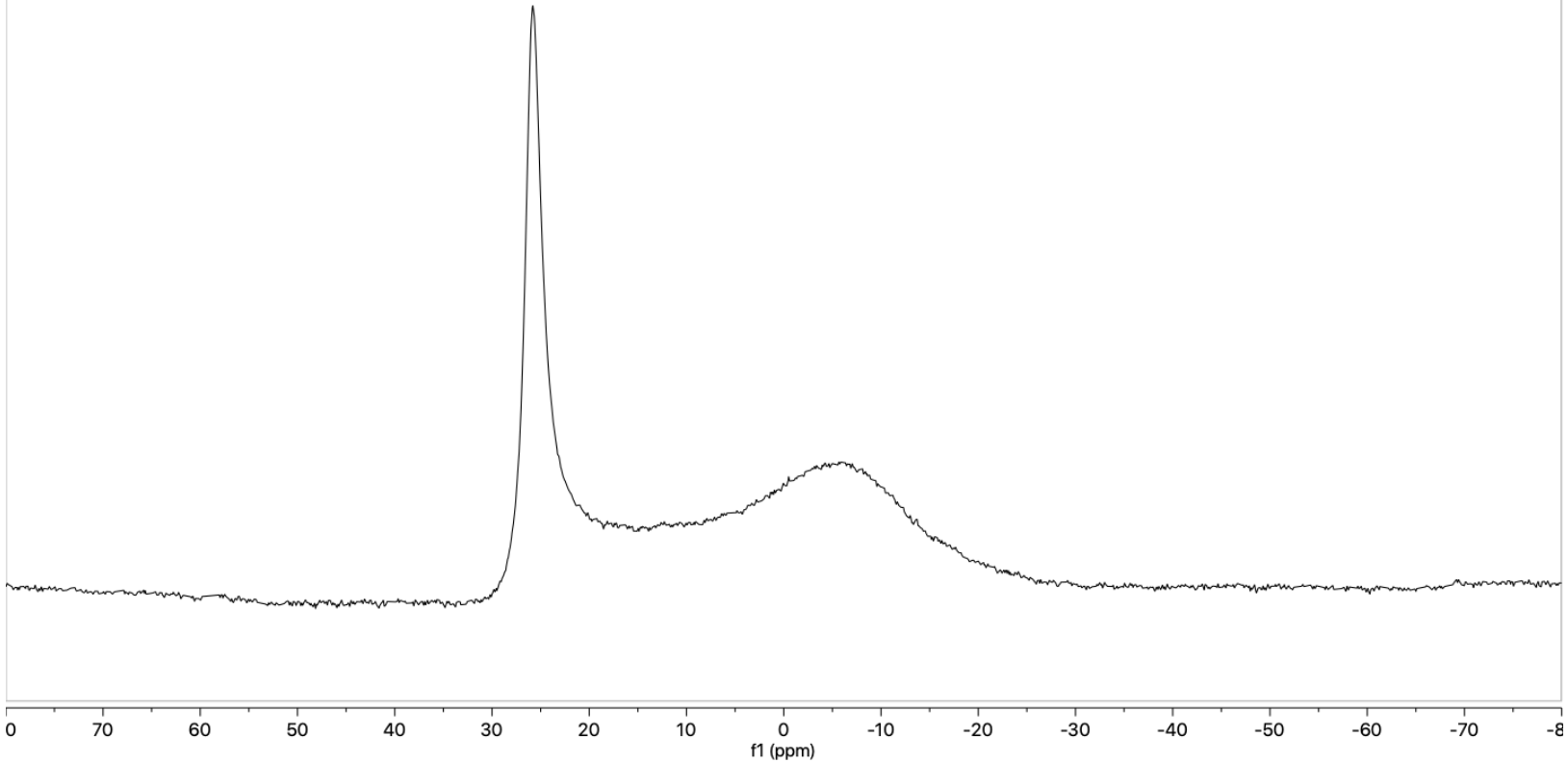
54.00 CD2Cl2



	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	0.1000
4	Acquisition Date	2016-10-16T12:07:46
5	Spectrometer Frequency	160.38
6	Nucleus	11B

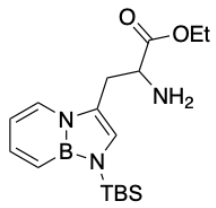


1.72



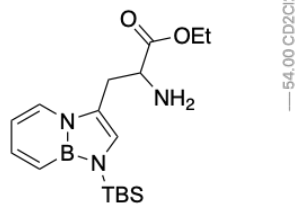
	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	5.0000
4	Acquisition Date	2015-05-27T14:11:13
5	Spectrometer Frequency	499.88
6	Nucleus	1H

5.32 CD2Cl2



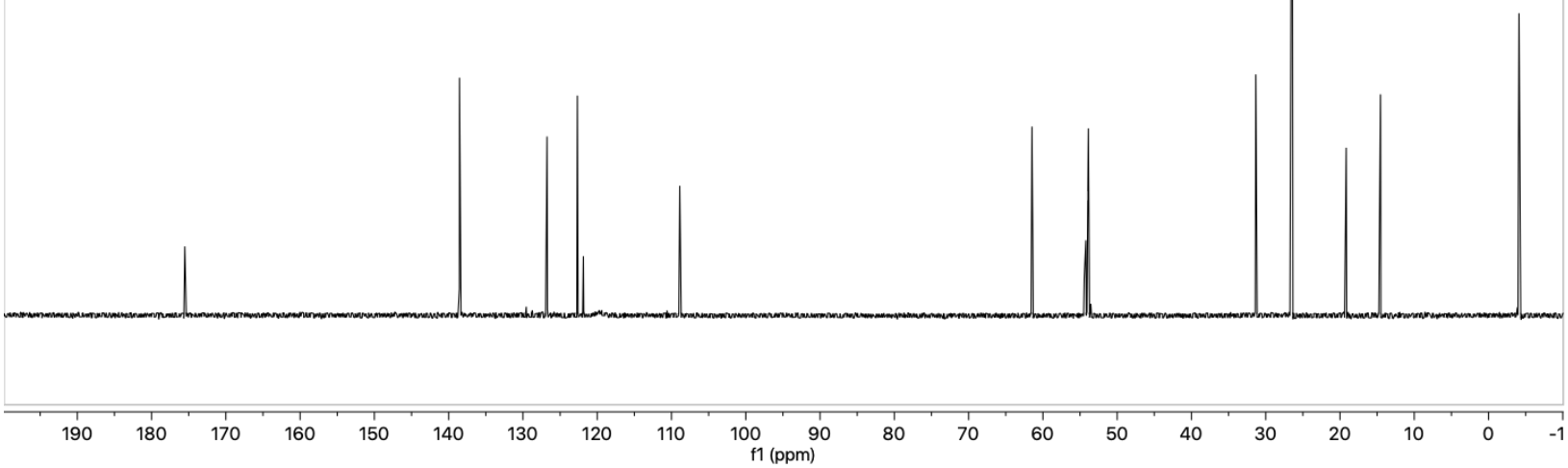
1.73

	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	1.0000
4	Acquisition Date	2015-05-27T14:16:05
5	Spectrometer Frequency	125.71
6	Nucleus	13C

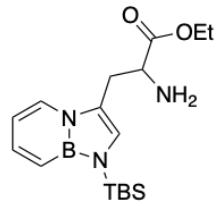


1.73

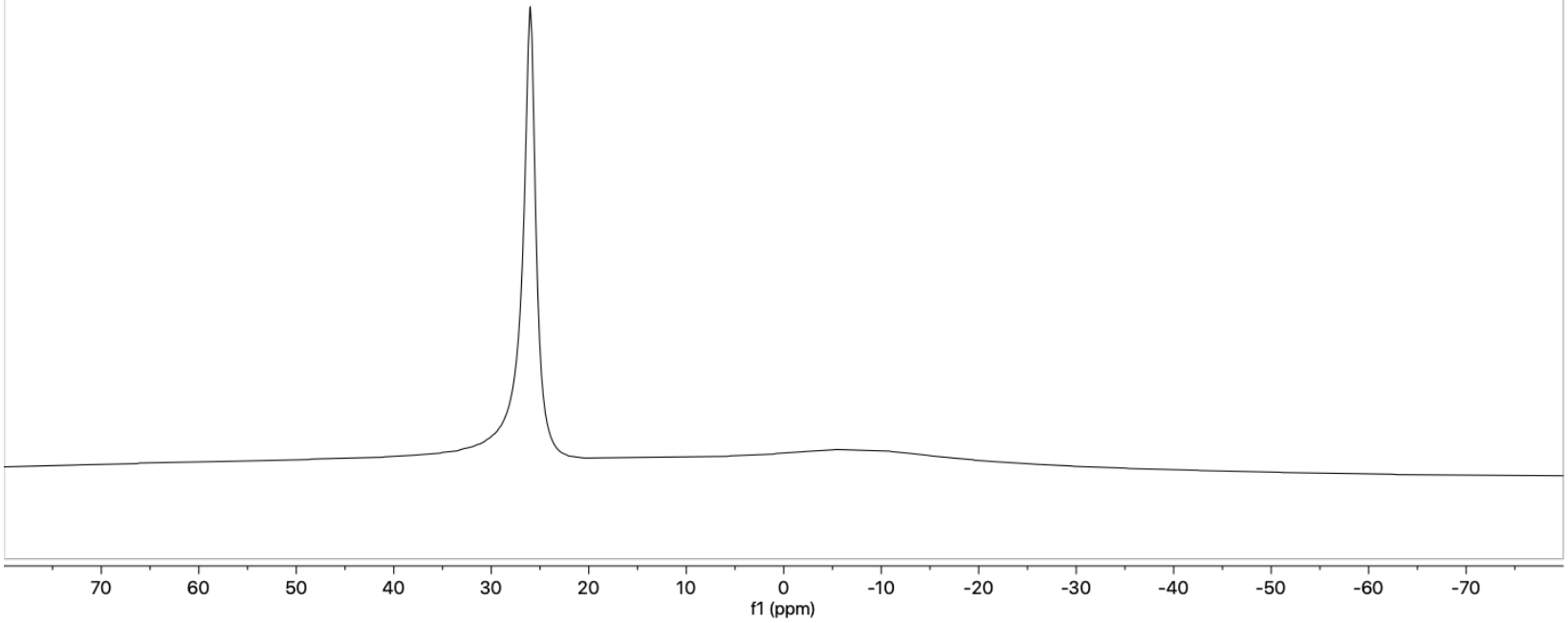
54.00 CD2Cl2



	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	0.1000
4	Acquisition Date	2015-05-27T14:06:45
5	Spectrometer Frequency	160.38
6	Nucleus	11B

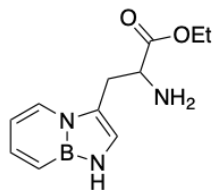


1.73

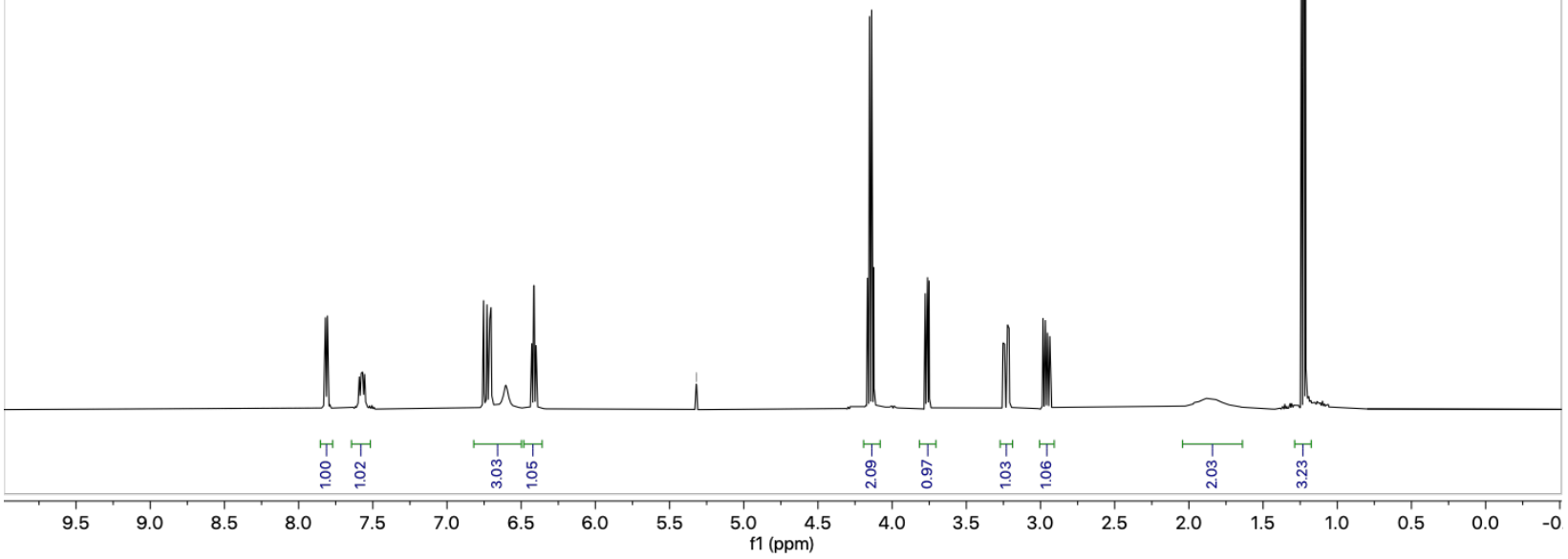


	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	2.0000
4	Acquisition Date	2015-08-25T12:39:49
5	Spectrometer Frequency	499.88
6	Nucleus	1H

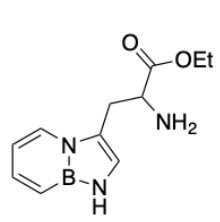
5.32 CD2Cl2



1.74

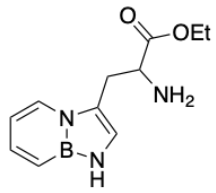


	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	1.0000
4	Acquisition Date	2015-08-25T12:45:54
5	Spectrometer Frequency	125.71
6	Nucleus	13C

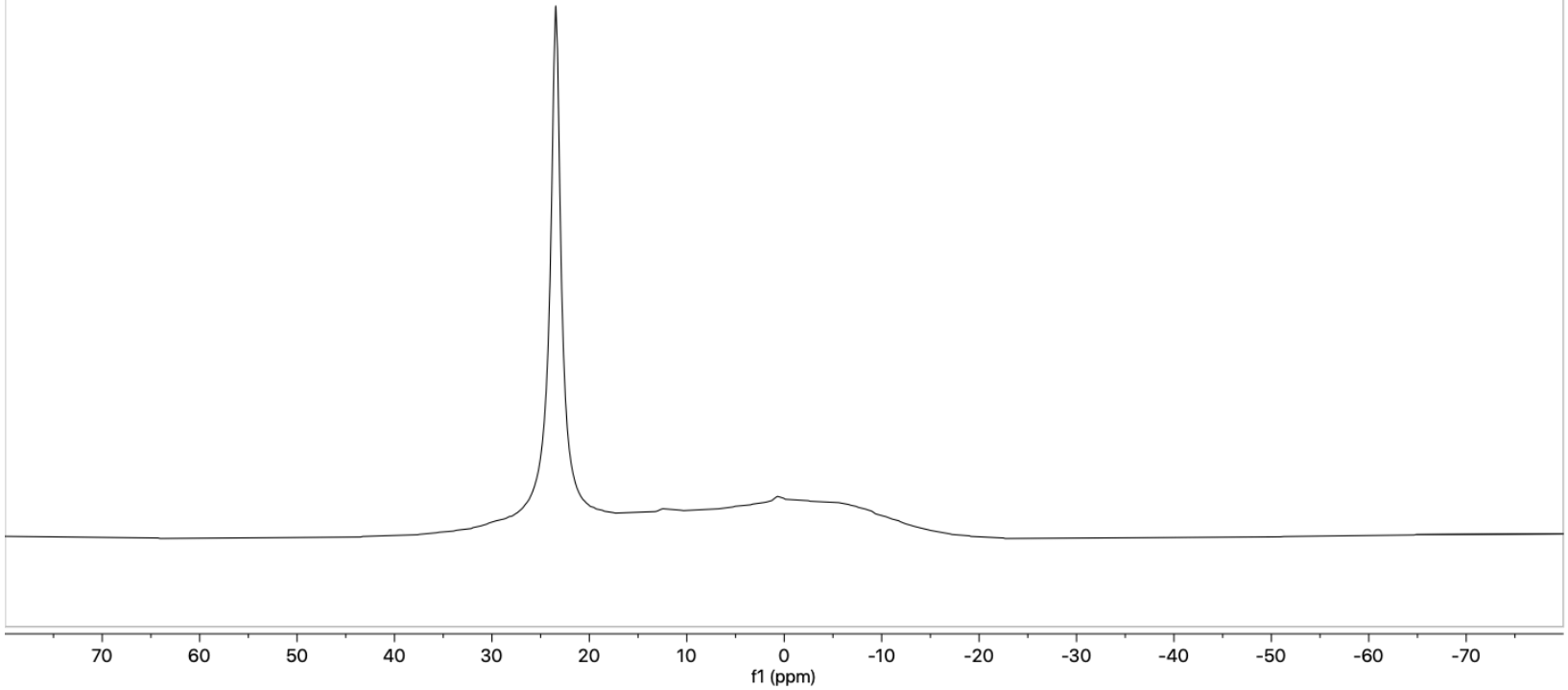


1.74

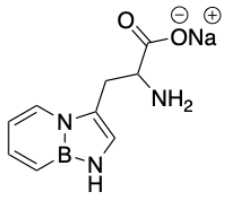
	Parameter	Value
1	Solvent	cd2cl2
2	Temperature	25.0
3	Relaxation Delay	0.1000
4	Acquisition Date	2015-08-25T12:35:22
5	Spectrometer Frequency	160.38
6	Nucleus	11B



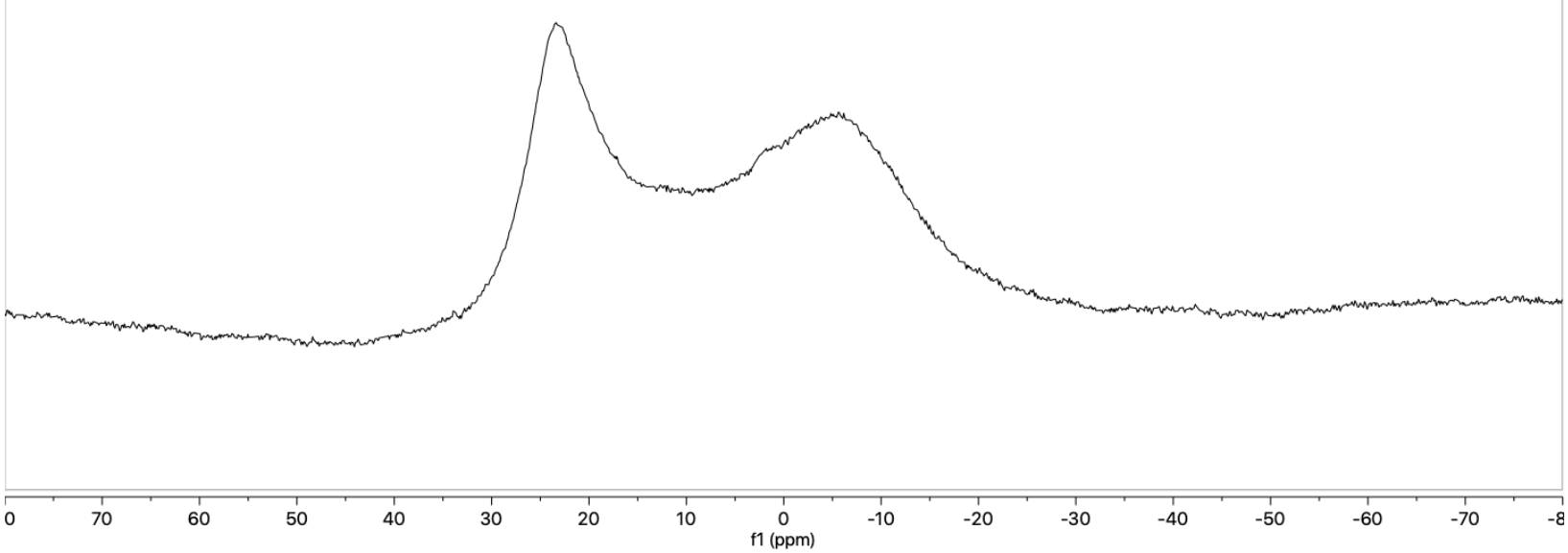
1.74



	Parameter	Value
1	Solvent	dmso
2	Temperature	25.0
3	Relaxation Delay	0.1000
4	Acquisition Date	2016-10-14T11:28:05
5	Spectrometer Frequency	160.38
6	Nucleus	11B



1.75

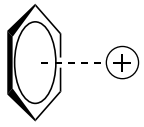


Chapter 2: Effect of BN/CC Isosterism on the Cation-π Binding Ability of Indole

2.1. Introduction: Noncovalent Forces Involving Arenes

Noncovalent inter- and intramolecular forces containing arenes play a key role in many chemical and biological processes.[1] These forces are defined as attractive, non-covalent interactions between two species and include arene-arene,[2] sulfur-π,[3] anion-π,[4] and cation-π[5] interactions as well as hydrogen-bonding to arenes.[6] One of the more prominent of these interactions is the cation-π interaction (Figure 1.1).

Figure 2.1 Representation of a cation-π interaction with benzene



The cation-π interaction is ubiquitous in many areas of biology and chemistry, including molecular recognition, materials chemistry, and catalysis, amongst many others.[7]

---

[1] For reviews, see: a) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Arenes Interactions with Aromatic Rings in Chemical and Biological Recognition *Angew. Chem. Int. Ed.* **2003**, *42*, 1210–1250. b) Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808–4842.

[2] a) Hunter, C. A.; Sanders, J. K. M. The Nature of π-π Interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534. b) Martinez, C. R.; Iverson, B. L. Rethinking the Term "Pi-Stacking." *Chem. Sci.* **2012**, *3*, 2191–2201.

[3] a) Reid, K. S. C.; Lindley, P. F.; Thornton, J. M. Sulphur-Aromatic Interactions in Proteins. *FEBS Lett.* **1985**, *190*, 209–213. b) Viguera, A. R.; Serrano, L. Side-Chain Interactions between Sulfur-Containing Amino Acids and Phenylalanine in α-Helices. *Biochemistry* **1995**, *34*, 8771–8779. c) Forbes, C. R.; Sinha, S. K.; Ganguly, H. K.; Bai, S.; Yap, G. P. A.; Patel, S.; Zondlo, N. J. Insights into Thiol-Aromatic Interactions: A Stereoelectronic Basis for S-H/π Interactions. *J. Am. Chem. Soc.* **2017**, *139*, 1842–1855.

[4] a) Schottel, B. L.; Chifotides, H. T.; Dunbar, K. R. Anion-π Interactions. *Chem. Soc. Rev.* **2008**, *37*, 68–83. b) Frontera, A.; Gamez, P.; Mascal, M.; Mooibroek, T. J.; Reedijk, J. Putting Anion-π Interactions into Perspective. *Angew. Chem. Int. Ed.* **2011**, *50*, 9564–9583. c) Giese, M.; Albrecht, M.; Rissanen, K. Anion-π Interactions with Fluoroarenes. *Chem. Rev.* **2015**, *115*, 8867–8895.

[5] Ma, J. C.; Dougherty, D. A. The Cation-π Interaction. *Chem. Rev.* **1997**, *97*, 1303–1324.

[6] a) Levitt, M.; Perutz, M. F. Aromatic Rings Act as Hydrogen Bond Acceptors. *J. Mol. Biol.* **1988**, *201*, 751–754. b) Steiner, T.; Koellner, G. Hydrogen Bonds with π-Acceptors in Proteins: Frequencies and Role in Stabilizing Local 3D Structures. *J. Mol. Biol.* **2001**, *305*, 535–557.

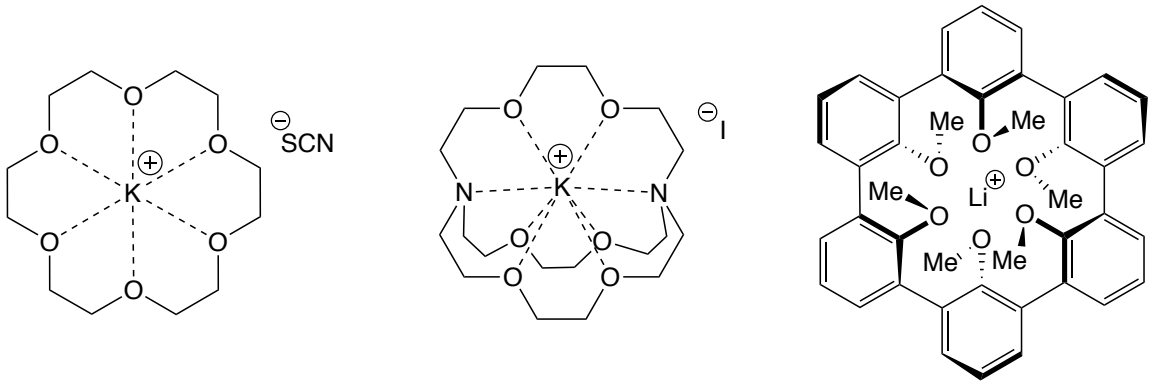
[7] For a review, see: Mahadevi, A. S.; Sastry, G. N. Cation−π Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chem. Rev.* **2013**, *113*, 2100–2138.

Since its discovery in the 1980's, a lot of work has been done to analyze the characteristics and roles of this interaction.

2.2. Introduction: Supramolecular Chemistry and Molecular Recognition

The field of supramolecular chemistry is the chemistry of molecular assemblies and the intermolecular bonds that hold molecules together, or as Lehn defines it, "chemistry beyond the molecule."[8] This field began with the binding of cations to crown ethers and cryptands, which led to the study of even more complex hosts (Figure 2.2). The pioneers of this chemistry – Pedersen, Lehn, and Cram – were awarded the Nobel Prize in Chemistry in 1987 for their work in this area.[9]

Figure 2.2 Pioneering Work in Supramolecular Chemistry



Crown-ether complex
Pedersen

Cryptand complex
Lehn

Host-guest complex
Cram

Molecular recognition is an important aspect of supramolecular chemistry and is the specific and non-covalent association between a host molecule, or receptor, and a guest substrate, or ligand.[10] Key players in molecular recognition are the intermolecular forces that form attractive interactions between the associating molecules. Many biological

[8] Lehn, J.-M. Supramolecular Chemistry. *Science.* **1993**, *260*, 1762–1763.
[9] Cram, D. J. The Design of Molecular Hosts, Guests, and Their Complexes *Angew. Chem. Int. Ed.* **1988**, *27*, 1009–1112. b) Pedersen, C. J. The Discovery of Crown Ethers (Noble Lecture). *Angew. Chem. Int. Ed.* **1988**, *27*, 1021–1027. c) Lehn, J. Supramolecular Chemistry-Scope and Perspectives. *Chem. Scr.* **1988**, *23*, 237.
[10] Lehn, J. Perspectives in Supramolecular Chemistry - From Molecular Recognition towards Molecular Information Processing and Self-Organization. *Angew. Chem. Int. Ed.* **1990**, *29*, 1304–1319.

processes rely on molecular recognition, such as enzyme-substrate binding, binding of neurotransmitters to neuroreceptors, ion binding to potassium ion-channels, and protein-protein interactions, amongst many other processes.

2.3. Background: Cation-π Interactions

2.3.1. The Discovery of the Cation-π Interaction: Gas Phase Studies

The cation-π interaction was first reported by Kebarle and coworkers in 1981.[11] In this work, they observed the binding interaction of a potassium ion (K^+) with benzene in the gas phase and measured it to have a ΔH = 19.2 kcal/mol (Figure 2.3). Surprisingly, they found K^+ bound preferentially with benzene over water: the K^+-water interaction was found to be 17.9 kcal/mol. Since benzene has no permanent dipole as water does, this warranted an in-depth analysis. Through calculations, they discovered the most stable conformation was with the K^+ ion situated over the center of the ring, which suggested it was interacting with the arene's quadrupole.

Figure 2.3 Interaction of K^+ with benzene and water and associated binding enthapies

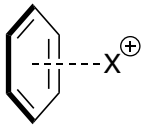
19.2 kcal/mol **17.9 kcal/mol**

A few years later, Meot-Ner and Deakyne demonstrated that a more complex cation, ammonium (NH_4^+), bound to olefins and benzene derivatives.[12] They measured the

[11] Sunner, J.; Nishizawa, K.; Kebarle, P. Ion-Solvent Molecule Interactions in the Gas Phase. The Potassium Ion and Benzene. *J. Phys. Chem.* **1981**, *85*, 1814–1820.
[12] Deakyne, C. A.; Meot-Ner, M. N. Unconventional Ionic Hydrogen Bonds. 2. NH+•••π. Complexes of Onium Ions with Olefins and Benzene Derivatives. *J. Am. Chem. Soc.* **1985**, *107*, 474–479.

NH_4^+-benzene interaction as having a binding energy of ΔH = 19.3 kcal/mol, nearly the same as that of K^+ and benzene (Table 2.1). It was also shown that the ammonium interaction decreased with fluorinated arenes, predictable due to the decreased quadrupole. Another publication indicated that increasing the size of the ammonium cation to tetramethylammonium (NMe_4^+) results in a decrease of binding energy to 9.4 kcal/mol (Table 2.1).[13]

Table 2.1 Binding enthalpies of various cations and benzene



Cation (X^+)	Li^+	Na^+	K^+	Rb^+	Cs^+	NH_4^+	NMe_4^+
Binding Enthalpies (kcal/mol)*	38.3	28.0	19.2	15.0	14.1	19.3	9.4

*Gas phase measurements

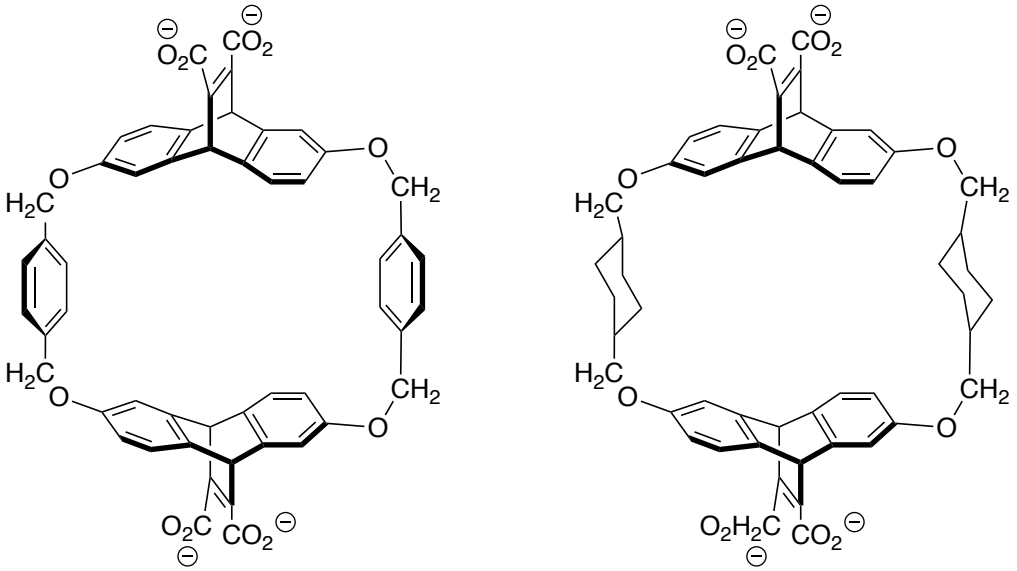
Since Kebarle's work, the rest of the series of mono-cationic alkali metals have been measured in the gas phase (Table 2.1). Guo and coworkers measured the Na^+-benzene interaction to have a ΔH = 28.0 kcal/mol.[14] In another work, Taft and coworkers measured the free energy of the Li^+-benzene interaction, which was later used to calculate the binding energy, which was 38.3 kcal/mol.[15,7] In a more recent work, Amicangelo and Armentrout

[13] Meot-Ner, M. N.; Deakyne, C. A. Unconventional Ionic Hydrogen Bonds. 1. CHδ+•••X. Complexes of Quaternary Ions with n- and π-Donors. *J. Am. Chem. Soc.* **1985**, *107*, 469–474.
[14] Guo, B. C.; Purnell, J. W.; Castleman, A. W. The Clustering Reactions of Benzene with Sodium and Lead Ions. *Chem. Phys. Lett.* **1990**, *168*, 155–160.
[15] Taft, R. W.; Anvia, F.; Gal, J. F.; Walsh, S.; Capon, M.; Holmes, M. C.; Hosn, K.; Oloumi, G.; Vasanwala, R.; Yazdani, S. Free Energies of Cation-Molecule Complex Formation and of Cation-Solvent Transfers. *Pure Appl. Chem.* **1990**, *62*, 17–23.

measured both the Rb^+-benzene and Cs^+-benzene interactions and found them to be 15.0 and 14.1 kcal/mol, respectively.[16]

2.3.2. The Emergence of the Cation-π Interactions: Solution Phase Studies

Figure 2.4 Cyclophane hosts developed by Dougherty and coworkers



Host 2.1 **Host 2.2**

Dougherty and coworkers pioneered the work on exploring cation-π interactions in solution.[17] They demonstrated that the binding of cationic (or neutral) guests in cyclophane hosts **2.1** and **2.2** in organic and aqueous media was largely due to the interaction of the electron-rich host with the electron-deficient guest (Figure 2.4).[18] Through a series of

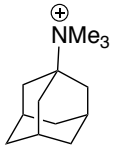
[16] Amicangelo, J. C.; Armentrout, P. B. Absolute Binding Energies of Alkali-Metal Cation Complexes with Benzene Determined by Threshold Collision-Induced Dissociation Experiments and Ab Initio Theory. *J. Phys. Chem. A* **2000**, *104*, 11420–11432.

[17] Dougherty, D. A. The Cation-π Interaction. *Acc. Chem. Res.* **2013**, *46*, 885–893.

[18] a) Shepodd, T. J.; Petti, M. A.; Dougherty, D. A. Tight, Oriented Binding of an Aliphatic Guest by a New Class of Water-Soluble Molecules with Hydrophobic Binding Sites. *J. Am. Chem. Soc.* **1986**, *108*, 6085–6087. b) Shepodd, T. J.; Petti, M. A.; Dougherty, D. A. Molecular Recognition in Aqueous Media: Donor-Acceptor and Ion-Dipole Interactions Produce Tight Binding for Highly Soluble Guests. *J. Am. Chem. Soc.* **1988**, *110*, 1983–1985. c) Barrans, R. E.; Dougherty, D. A.; Petti, M. A.; Shepodd, T. J. "Hydrophobic" Binding of Water-Soluble Guests by High-Symmetry, Chiral Hosts. An Electron-Rich Receptor Site with a General Affinity for Quaternary Ammonium Compounds and Electron-Deficient π Systems. *J. Am. Chem. Soc.* **1988**, *110*, 6825–6840. d) Stauffer, D. A.; Dougherty, D. A. Ion-Dipole Effect as a Force for Molecular Recognition in Organic Media. *Tetrahedron Lett.* **1988**, *29*, 6039–6042.

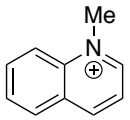
manipulations of the host and guest, they found the binding interactions were stronger for cationic guests and aromatic hosts (Figure 2.5). When they either replaced the guest with a neutral molecule or the arenes in the host with saturated cyclohexanes (Host **2.2**), they saw a significant decrease in binding.[19]

Figure 2.5 Binding free energies and dissociation constants for the interaction with **Host 2.1** in water



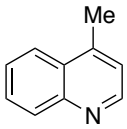
2.3

ATMA
$-\Delta G° = 6.7$ kcal/mol
$K_D = 12\ \mu M$



2.4

NMQ
$-\Delta G° = 8.4$ kcal/mol
$K_D = 600$ nM



2.5

4-Me-quinoline
$-\Delta G° = 5.9$ kcal/mol
$K_D = 47\ \mu M$

Schneider and coworkers were also able to identify a favorable interaction between cationic and aromatic species in multiple studies.[20] They demonstrated strong arene-ammonium interactions using their azacyclophane host **2.6** (Figure 2.6). In one particular case, they measured the binding of bicyclic guests **2.9**, **2.10**, and **2.11** and found that the binding decreased as the naphthalene became more saturated, which is consistent with a cation-π interaction being involved in the binding (Table 2.2).[21] Even more so, when using anionic hosts **2.7** and **2.8**, they observed a decrease in binding with the neutral guests **2.9** and **2.12**, which indicates that the cationic property of host **2.6** was important. In these

[19] Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; McCurdy, A.; Dougherty, D. A. Molecular Recognition in Aqueous Media. New Binding Studies Provide Further Insights into the Cation-π Interaction and Related Phenomena. *J. Am. Chem. Soc.* **1993**, *115*, 9907–9919.

[20] a) Schneider, H.-J.; Blatter, T.; Zimmermann, P. Host-Guest Complexes with Closed, Half-Open, and Stretched Receptors: Hydrophobic Cavity Effects and Induced Pole-Dipole Interactions. *Angew. Chem. Int. Ed.* **1990**, *29*, 1161–1162. b) Schneider, H.-J; Güttes, D.; Schneider, U. A Macrobicyclic Polyphenoxide as Receptor Analogue for Choline and Related Ammonium Compounds. *Angew. Chem. Int. Ed.***1986**, *25*, 647–649.

[21] Schneider, H.-J; Blatter, T. Modification of Hydrophobic and Polar Interactions by Charged Groups in Synthetic Host-Guest Complexes. *Angew. Chem. Int. Ed* **1988**, *27*, 1163–1164. b) Schneider, H.-J.; Blatter, T.; Simova, S.; Theis, I. Large Binding Constant Differences between Aromatic and Aliphatic Substrates in Positively Charged Cavities Indicative of Higher Order Electric Effects. *J. Chem. Soc. Chem. Commun.* **1989**, *9*, 580–581.

studies, Schneider was able to determine that each phenyl ring in the cyclophane hosts contributed ~0.5 kcal/mol in binding energy due to the cation-π interaction.[22]

Figure 2.6 Hosts and guests studied by Schneider and coworkers

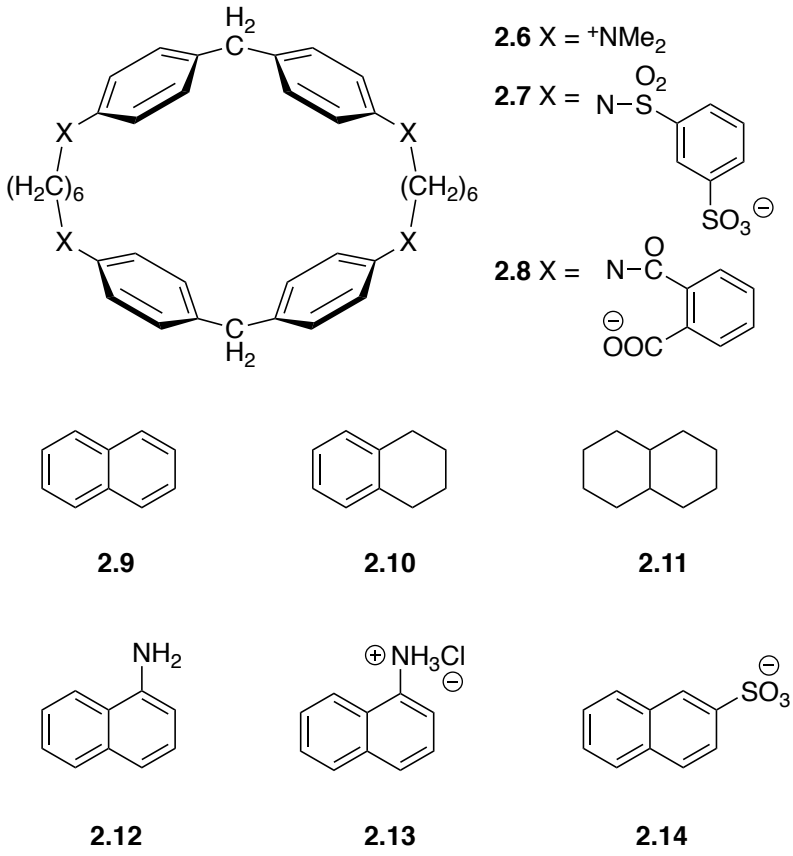


Table 2.2 Binding free enthalpies (kcal/mol)

	2.9	**2.10**	**2.11**	**2.12**	**2.13**	**2.14**
2.6	4.0	3.1	1.6	3.8	1.1	5.4
2.7	3.1	—	—	3.5	3.8	2.1
2.8	3.2	—	—	2.7	—	1.7

[22] Schneider, H.-J.; Schiestel, T.; Zimmermann, P. The Incremental Approach to Noncovalent Interactions: Coulomb and van Der Waals Effects in Organic Ion Pairs. *J. Am. Chem. Soc.* **1992**, *114*, 7698–7703.

One important factor to consider when characterizing cation-π interactions in solution phase studies is the magnitude of the desolvation energy of the guest.[18b] It is well known that cations are very strongly solvated in water and polar solvents and in order for a cation to bind in a host, it must overcome that desolvation energy to have a favorable interaction. In fact, in the earlier aqueous phase cation-binding studies,[18a] Dougherty used quaternary ammonium compounds that were soluble in both organic and aqueous solvents to better study the interaction since simple cations would be too soluble in water.

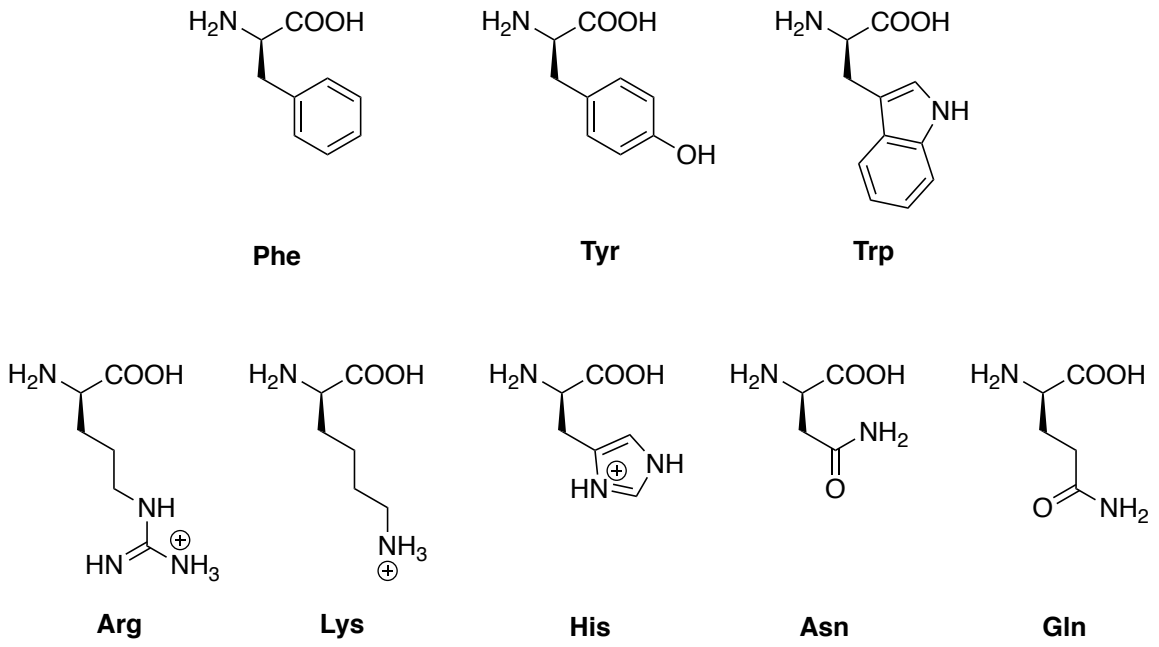
2.3.3. Cation-π Interactions in Protein Secondary Structure

Evidence of amino groups interacting with aromatic amino acids in biological systems was first reported by Burley and Petsko in 1986.[23] They examined 33 protein crystal structures for the proximity of aromatic amino acid side chains (Phe, Tyr, and Trp) with amine-containing amino acid side chains (Arg, Lys, Asn, Gln, and His) and found that the positively charged amino groups in contact with arenes were within 6Å of the arene ring-center (Figure 2.7). Later, however, Thornton reported that in these amino-aromatic interactions, the NH proton interacting with the face of the aromatic ring (i.e. an electrostatic interaction of the positively polarized protons with the electron density of the aromatic ring) are actually rare: in most cases, the sp^2 NH near an aromatic residue is stacked with the ring (i.e. a π-stacking interaction) and participating in hydrogen-bonding interactions with other amino acids (Figure 2.8).[24] Despite this discovery, Burley and Petsko's work alerted structural biologists to the potential for aromatics to participate in polar-π interactions.

---

[23] Burley, S. K.; Petsko, G. A. Amino-Aromatic Interactions in Proteins. *FEBS Lett.* **1986**, *203*, 139–143.
[24] Mitchell, J. B. O.; Nandi, C. L.; McDonald, I. K.; Thornton, J. M.; Price, S. L. Amino/Aromatic Interactions in Proteins: Is the Evidence Stacked against Hydrogen Bonding? *J. Mol. Biol*. **1994**, *239*, 315–331.

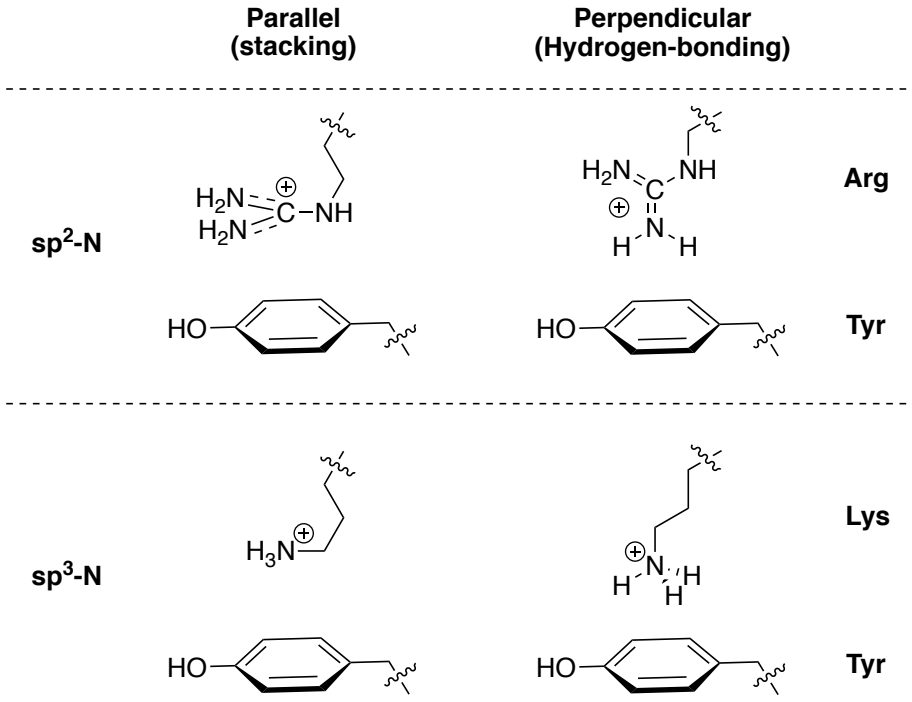
Figure 2.7 Amino acids studied in protein structure analyses



Years later in 1999, Gallivan and Dougherty revisited the topic of analyzing the role cation-π interactions play in protein structures.[25] Using energetics-based criteria in their statistical analysis (as opposed to the geometry-based criteria of earlier studies),[23] they analyzed 593 high resolution, nonhomologous protein structures. In their analysis, they considered only cationic and aromatic side-chains (Arg or Lys interacting with Phe, Tyr, or Trp) and evaluated all possible cation-π interactions for those that made a significant energetic contribution to the stability of the proteins. In summary, they found that there is 1 cation-π interaction for every 77 residues of protein length in the PDB, over 500,000 cation-π interactions were in the PDB at the time of publication, and 26% of all Trp residues experience a significant cation-π interaction with Arg or Lys in proteins.

[25]Gallivan, J. P.; Dougherty, D. A. Cation-π Interactions in Structural Biology. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459–9464.

Figure 2.8 Representative examples of geometries of amino-aromatic interactions



Many studies of protein structures from the PDB have since emerged.[26,27] In one case, Crowley reported that cation-π interactions are abundant at protein-protein interfaces.[28] They found that about half of all protein complexes and a third of protein homodimers exhibited at least 1 cation-π interaction and that Arg-Tyr interactions were the most common. With the ever-growing PDB and similar databases, there are sure to be more of these studies emerging to further explore the role of cation-π interactions in structural biology.

[26] a) Hou, Q.; Bourgeas, R.; Pucci, F.; Rooman, M. Computational Analysis of the Amino Acid Interactions That Promote or Decrease Protein Solubility. *Sci. Rep.* **2018**, *8*, 1–13. b) Kumar, K.; Woo, S. M.; Siu, T.; Cortopassi, W. A.; Duarte, F.; Paton, R. S. Cation-π Interactions in Protein-Ligand Binding: Theory and Data-Mining Reveal Different Roles for Lysine and Arginine. *Chem. Sci.* **2018**, *9*, 2655–2665. c) Mbaye, M. N.; Hou, Q.; Basu, S.; Teheux, F.; Pucci, F.; Rooman, M. A Comprehensive Computational Study of Amino Acid Interactions in Membrane Proteins. *Sci. Rep.* **2019**, *9*, 1–14.

[27] For more references, see: Ma, J. C.; Dougherty, D. A. The Cation-π Interaction. *Chem. Rev.* **1997**, *97*, 1303–1324.

[28] Crowley, P. B.; Golovin, A. Cation-π Interactions in Protein-Protein Interfaces. *Proteins Struct. Funct. Genet.* **2005**, *59*, 231–239.

With the awareness of the cation-π interaction in proteins, more efforts into determining its role in the protein's overall structure and function have emerged.[29] In many cases, simpler, model studies have been performed to elucidate how the cation-π interaction affects the proteins.[30] For instance, short peptides and mini-proteins have been used to simplify the targeted portion of the protein structure.[31]

In a recent example from Bonneau and Kirshenbaum, a 19-residue miniature protein that can be used to emulate the "R…WSXWS motif"[32] found in the fibronectin type III (Fn3) domain containing proteins was described.[33] This motif forms a cation-π network between an Arg and two Trp residues that stabilizes a rare class of secondary structure, the left-handed poly proline type-II helix (PPII), in proteins and has been observed in many X-Ray protein crystal structures. In this work, they were able to demonstrate that they could replicate the cation-π network from a larger protein in a relatively small mini-protein.

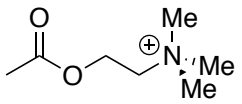
2.3.4. Cation-π Interactions in Ligand Binding and Ion Channels

While many of Dougherty's earlier works focused on the binding of cationic heterocycles and bulky, hydrophobic ammoniums, some of his later reports switched focus to more biologically relevant molecules. In 1990, they reported the binding of acetylcholine (ACh), a neurotransmitter, in their cyclophane hosts (Figure 2.9).[34] ACh binding sites have

[29] Mahadevi, A. S.; Sastry, G. N. Cation−π Interaction: Its Role and Relevance in Chemistry, Biology, and Material Science. *Chem. Rev.* **2013**, *113*, 2100–2138.
[30] Kool, E. T.; Waters, M. L. The Model Student: What Chemical Model Systems Can Teach Us about Biology. *Nat. Chem. Biol.* **2007**, *3*, 70–73.
[31] Baker, E. G.; Bartlett, G. J.; Porter Goff, K. L.; Woolfson, D. N. Miniprotein Design: Past, Present, and Prospects. *Acc. Chem. Res.* **2017**, *50*, 2085–2092.
[32] R = Arginine, W = Tryptophan, S = Serine, and X represents a variable amino acid.
[33] Craven, T. W.; Cho, M. K.; Traaseth, N. J.; Bonneau, R.; Kirshenbaum, K. A Miniature Protein Stabilized by a Cation-π Interaction Network. *J. Am. Chem. Soc.* **2016**, *138*, 1543–1550.
[34] Dougherty, D. A.; Stauffer, D. A. Acetylcholine Binding by a Synthetic Receptor: Implications for Biological Recognition. *Science* **1990**, *250*, 1558–1560.

been shown to contain multiple aromatic amino acid residues. For instance, the molecular structure of ACh esterase (AChE) - a natural ACh binding site that plays a crucial role in cholinergic synapses and terminating synaptic transmission by hydrolyzing ACh to choline and acetate - contains 14 aromatic amino acid residues that make up a large portion of the ACh-binding site and a ~20 Å cleft that leads from the surface of the enzyme to the active site (known as the "aromatic gorge").[35]

Figure 2.9 Structure of Acetylcholine



2.15

ACh

The nicotinic ACh receptor (nAChR) is another important and well-studied neuroreceptor that binds ACh and has been a source of evidence of cation-π binding interactions in ACh binding sites (Figure 2.10).[36] The nAChR is a ligand-gated ion channel in which ACh binds to the receptor, triggering a conformational change that opens an ion channel that is within the protein. Interest in this particular protein lies in that nicotine, a stimulant and cholinomimetic drug present in tobacco, is an agonist for this receptor and that nAChR plays an important role in nicotine addiction.[37] Many studies have been done to elucidate the role of cation-π interactions in the binding of ACh or nicotine.[7]

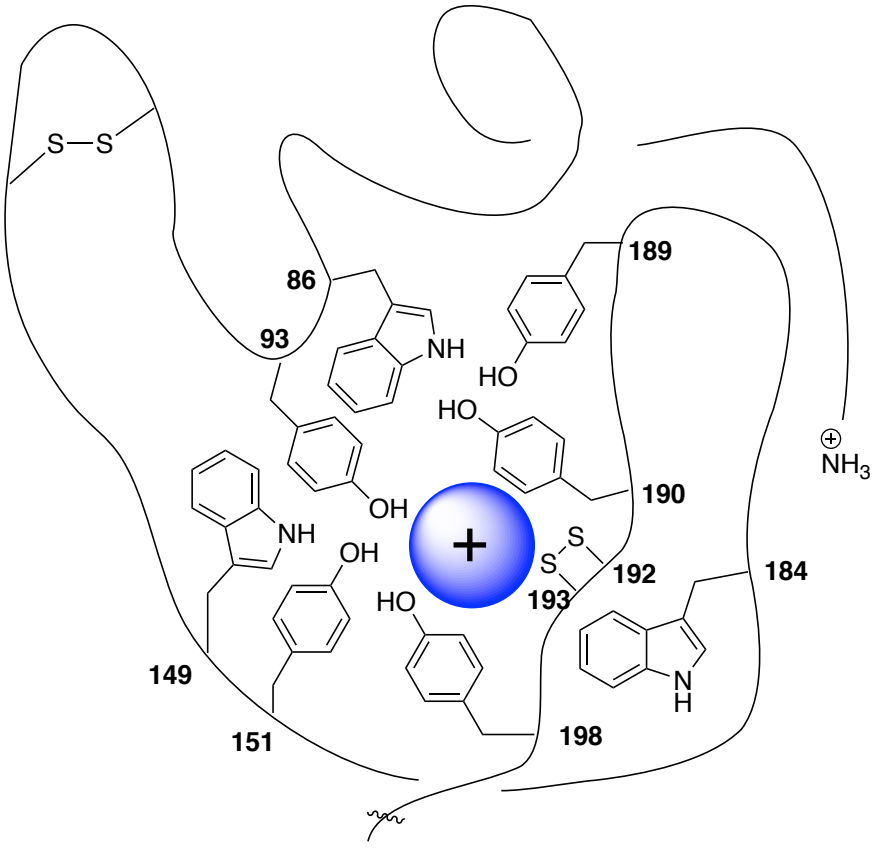
---

[35] Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. Atomic Structure of Acetylcholinesterase from Torpedo Californica: A Prototypic Acetylcholine-Binding Protein. *Science* **1991**, *253*, 872–879.

[36] Schmitt, J. D.; Sharples, C. G. V.; Caldwell, W. S. Molecular Recognition in Nicotinic Acetylcholine Receptors: The Importance of π-Cation Interactions. *J. Med. Chem.* **1999**, *42*, 3066–3074.

[37] Tapper, A. R.; McKinney, S. L.; Nashmi, R.; Schwarz, J.; Deshpande, P.; Labarca, C.; Whiteaker, P.; Marks, M. J.; Collins, A. C.; Lester, H. A. Nicotine Activation of α4* Receptors: Sufficient for Reward, Tolerance, and Sensitization. *Science.* **2004**, *306*, 1029–1032.

Figure 2.10 nAChR Agonist Binding Site



Aromatic amino acids populate many cation binding sites, in fact, a common motif seen in cationic binding sites is known as the "aromatic box".[17] One of the best examples of this is seen in the ACh binding proteins (AChBP).[38] The AChBP showed that the agonist binding site contains three tyrosine and two tryptophan residues that form an aromatic box. A similar motif has been found in multiple other proteins, including some completely unrelated proteins, which demonstrates the overarching importance of cation-π interactions in ligand recognition.[39,1b,7]

---

[38] Brejc, K.; Van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; Van Der Oost, J.; Smit, A. B.; Sixma, T. K. Crystal Structure of an ACh-Binding Protein Reveals the Ligand-Binding Domain of Nicotinic Receptors. *Nature* **2001**, *411*, 269–276.

[39]a) Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; et al. Structure of the Human M2 Muscarinic Acetylcholine Receptor Bound to an Antagonist. *Nature* **2012**, *482*, 547–551. b) Schärer, K.; Morgenthaler, M.; Paulini, R.; Obst-Sander, U.;

Cation-π interactions also play a major role in ion selectivity in ion channels. Cloning and sequencing of the *Shaker* voltage-dependent K^+ ion channel revealed the structure of these channels, which led to various studies on how the pore region selects for certain cations.[40] In 1992, Heginbotham and MacKinnon suggested that cation-π interactions may be responsible for ion selectivity in their studies on the mechanism in which tetraethylammonium (TEA), a K^+ channel inhibitor, interacts with the binding site.[41] This proposal was further supported by Dougherty and coworkers' report in 1993.[42] Through computational methods, they determined the selectivity of ion (K^+>Rb^+>>Na^+>Li^+) binding to benzene in an aqueous environment to mimic cation selectivity seen in voltage-gated potassium channels. Many other studies have since been performed to establish the role of cation-π interactions in ion channels.[43]

2.3.5. Cation-π Interactions in Biological Systems Involving Tryptophan

Despite its rarity in proteins, studies have shown that tryptophan, with its electron-rich indole side-chain, is one of the main contributors to cation-π interactions in biological

---

Banner, D. W.; Schlatter, D.; Benz, J.; Stihle, M.; Diederich, F. Quantification of Cation-π Interactions in Protein-Ligand Complexes: Crystal-Structure Analysis of Factor Xa Bound to a Quaternary Ammonium Ion Ligand. *Angew. Chemie - Int. Ed.* **2005**, *44*, 4400–4404.

[40] a) Miller, C. Annus Mirabilis of Potassium Channels. *Science.* **1991**, *252*, 1092–1096. b) Doyle, D. A.; Cabral, J. M.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The Structure of the Potassium Channel: Molecular Basis of K+ Conduction and Selectivity. *Science* **1998**, *280*, 69–77.

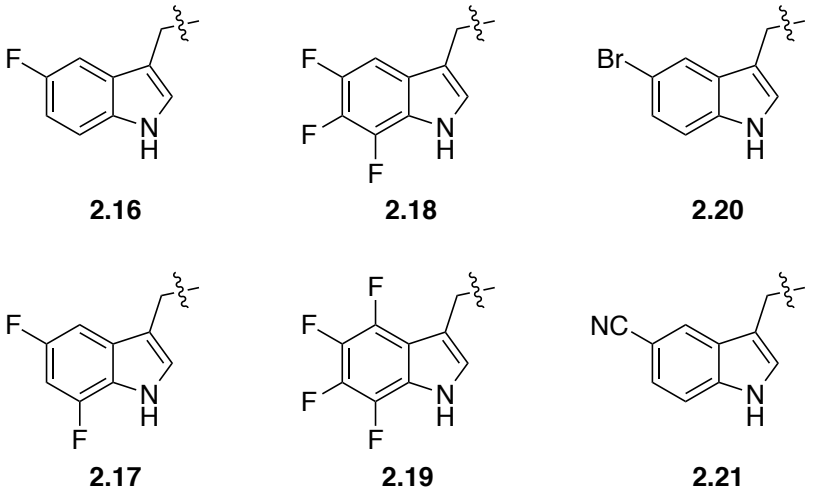
[41] Heginbotham, L.; MacKinnon, R. The Aromatic Binding Site for Tetraethylammonium Ion on Potassium Channels. *Neuron* **1992**, *8*, 483–491.

[42] Kumpf, R. A.; Dougherty, D. A. A Mechanism for Ion Selectivity in Potassium Channels: Computational Studies of Cation-π Interactions. *Science* **1993**, *261*, 1708–1710.

[43] a) Pless, S. A.; Galpin, J. D.; Niciforovic, A. P.; Ahern, C. A. Contributions of Counter-Charge in a Potassium Channel Voltage-Sensor Domain. *Nat. Chem. Biol.* **2011**, *7*, 617–623. b) Xu, Y.; Shin, H. G.; Szép, S.; Lu, Z. Physical Determinants of Strong Voltage Sensitivity of K + Channel Block. *Nat. Struct. Mol. Biol.* **2009**, *16*, 1252–1258. c) Fernandez, D.; Ghanta, A.; Kauffman, G. W.; Sanguinetti, M. C. Physicochemical Features of the HERG Channel Drug Binding Site. *J. Biol. Chem.* **2004**, *279*, 10120–10127. d) González, W.; Riedelsberger, J.; Morales-Navarro, S. E.; Caballero, J.; Alzate-Morales, J. H.; González-Nilo, F. D.; Dreyer, I. The pH Sensor of the Plant K +-Uptake Channel KAT1 Is Built from a Sensory Cloud Rather than from Single Key Amino Acids. *Biochem. J.* **2012**, *442*, 57–63. e) Ahern, C. A.; Eastwood, A. L.; Lester, H. A.; Dougherty, D. A.; Horn, R. A Cation-π Interaction between Extracellular TEA and an Aromatic Residue in Potassium Channels. *J. Gen. Physiol.* **2006**, *128*, 649–657.

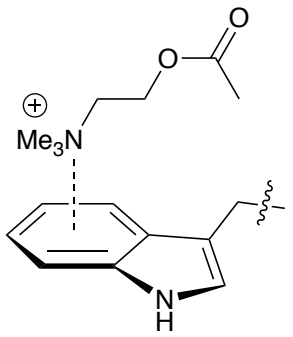
settings.[44] Of the three aromatic amino acids that participate in cation-π interactions, tryptophan is the strongest π-donor. Due to tryptophan's cation-π binding ability, it plays a crucial role in the stability of proteins. For instance, it has been shown to stabilize the secondary structure of β-hairpin peptides.[45]

Figure 2.11 Tryptophan Analogues Used in Binding Studies



Tryptophan is also prevalent in neurobiology and is found in neurotransmitter receptors, as shown in the early work of Dougherty, in which he reports his studies on tryptophan's role in the acetylcholine receptor in nAChR.[34] Dougherty and coworkers have demonstrated the importance of the tryptophan residue in the nAChR through the use of *in vivo* nonsense suppression method to incorporate tryptophan

Figure 2.12 Binding Model of Acetylcholine to Trp



[44] a) Dougherty, D. A. Cation-P Interactions Involving Aromatic Amino Acids. *J. Nutr.* **2007**, *137*, 1504–1508. b) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. Nicotine Binding to Brain Receptors Requires a Strong Cation- Interaction. *Nature* **2009**, *458*, 534–537.

[45] a) Riemen, A. J.; Waters, M. L. Design of Highly Stabilized B-Hairpin Peptides through Cation - π Interactions of Lysine and N-Methyllysine with an Aromatic Pocket. *Biochemistry* **2009**, *48*, 1525–1531. b) Santiveri, C. M.; Jiménez, M. A. Tryptophan Residues: Scarce in Proteins but Strong Stabilizers of β-Hairpin Peptides. *Biopolymers* **2010**, *94*, 779–790. c) Mahalakshmi, R. Aromatic Interactions in β-Hairpin Scaffold Stability: A Historical Perspective. *Arch. Biochem. Biophys.* **2019**, *661*, 39–49.

analogues with varying electronic characteristics (Figure 2.11).[46] This study helped them identify one Trp residue in the binding site that is responsible for the binding of ACh and other quaternary ammonium ligands (Figure 2.10, Trp149). (A model of ACh binding to trp can be seen in Figure 2.12). Progressive fluorination led to an increase in EC_{50} of acetylcholine and a decrease in the calculated binding energy, which indicates that tryptophan plays a key role in binding of ACh (Table 2.3).

Table 2.3 Binding Energies and EC_{50} Values of ACh Binding

Side Chain	Cation-π binding (kcal/mol)*	EC_{50} (μM)
Trp	32.6	1.2
2.16	27.5	4.7
2.17	23.3	13
2.18	18.9	34
2.19	14.4	65
2.20	27.8	2.0
2.21	21.5	114

* Calculated negative binding energies of a generic Na^+ to appropriate arene.

A subsequent study explored the role of Trp in the binding of various ammonium agonists to the nAChR and serotonergic (5-HT_{3A}) receptors (Figure 2.13).[47] Here they saw the same trend in decrease in binding strength with progressive fluorination. The energetic

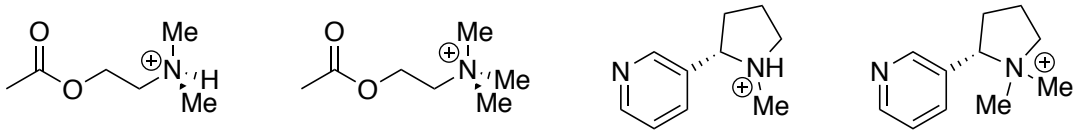
[46] Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A. From Ab Initio Quantum Mechanics to Molecular Neurobiology: A Cation-π Binding Site in the Nicotinic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12088–12093.

[47] Beene, D. L.; Brandt, G. S.; Zhong, W.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A. Cation-π Interactions in Ligand Recognition by Serotonergic (5-HT3A) and Nicotinic Acetylcholine Receptors: The Anomalous Binding Properties of Nicotine. *Biochemistry* **2002**, *41*, 10262–10269.

contribution of the cation-π interaction was estimated to be ~4 kcal/mol in the binding of serotonin to 5-HT_{3A}R and ~2 kcal/mol for ACh binding to nAChR.

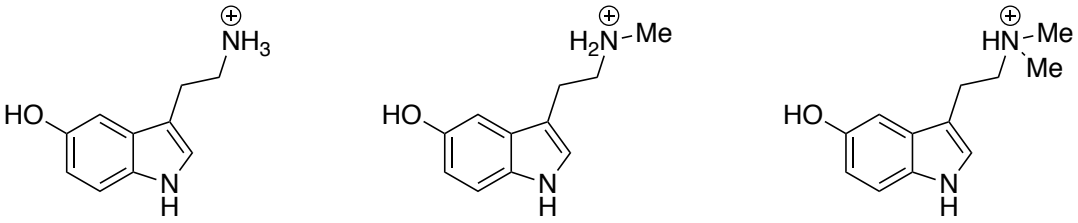
Figure 2.13 Agonists Explored in Binding to nAChR and 5-HT_{5A}R

nAChR binders



2.22 **2.15** **2.23** **2.24**

5-HT_{5A}R binders



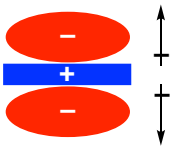
2.25 **2.26** **2.27**

2.3.6. The Fundamental Nature of the Cation-π Interaction

The cation-π interaction is defined as the interaction of a cationic species with an electron π-system and involves the interaction of a cation with a permanent quadrupole moment (Figure 2.14).[48] Since the discovery of the cation-π interaction and its importance in biological processes, theoretical and experimental studies have been done to determine its fundamental nature.[1b] Even so, there is still much debate on the full nature of the cation-π interactions. However, in many cases, a simplified model can be useful for predicting interactions between various cations and π-systems.

[48] Dougherty. D. A. Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr and Trp. *Science* **1996**, *271*, 163–169.

Figure 2.14 Schematic of the Quadrupole Moment of Benzene



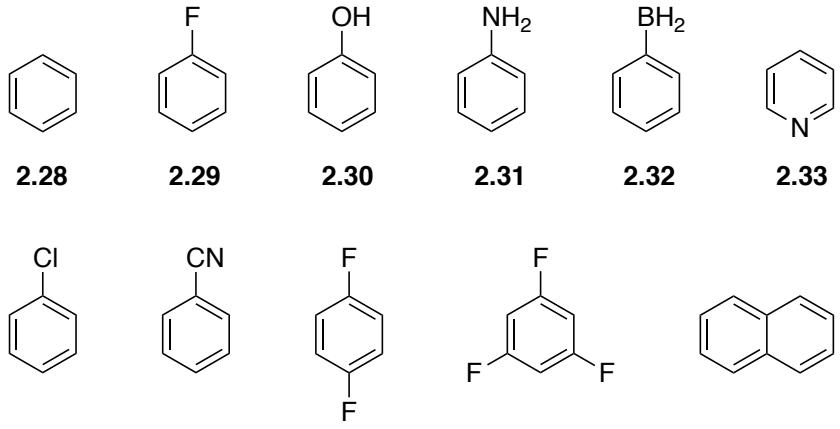
The electrostatic model, first proposed by Kebarle, is the generally accepted defining nature of the cation-π interaction.[49] The electrostatic component of the interaction was proposed to be the main contributing factor to the interaction partly due to the trends observed in gas phase studies. The comparison of simple alkali metal cations binding to benzene shows a classic electrostatic trend in binding strengths: $Li^+ > Na^+ > K^+ > Rb^+$. Smaller cations are expected to bind stronger in an electrostatic sequence, whereas if polarizability or dispersion forces were prominent, an opposite trend is expected.

Calculations have further shown that one of the main contributing factors to the cation-π interaction is the electrostatic component. Early work by Dougherty demonstrated the importance of the electrostatic component in the cation-π interaction through computational studies with the model system of binding Na^+ to various aromatic molecules that emphasized the electrostatic component of the π-system (Figure 2.15).[50] He showed that electrostatics can be used as a qualitative guide to cation-π interactions, and later, published a more extensive study on using electrostatic potential maps on biologically relevant aromatics to predict cation-π interactions.[51]

---

[49] Kebarle, P. Ion Thermochemistry and Solvation from Gas Phase Ion Equilibria. *Annu. Rev. Phys. Chem.* **1977**, *28*, 445–476.
[50] Mecozzi, S.; West, A. P.; Dougherty, D. A. Cation-π Interactions in Simple Aromatics: Electrostatics Provide a Predictive Tool. *J. Am. Chem. Soc.* **1996**, *118*, 2307–2308.
[51] Mecozzi, S.; West, A. P.; Dougherty, D. A. Cation-π Interactions in Aromatics of Biological and Medicinal Interest: Electrostatic Potential Surfaces as a Useful Qualitative Guide. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 10566–10571.

Figure 2.15 Arenes Explored in Computational Studies of Binding to Na^+



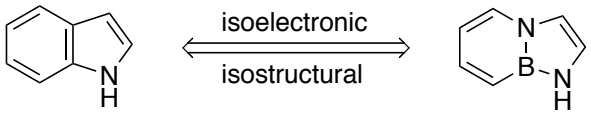
2.28 **2.29** **2.30** **2.31** **2.32** **2.33**

2.34 **2.35** **2.36** **2.37** **2.9**

2.4. Background: BN/CC Isosterism and Electronic Structure

2.4.1. Electronic Properties of BN-Indole

Figure 2.16 BN/CC Isosterism of Indole



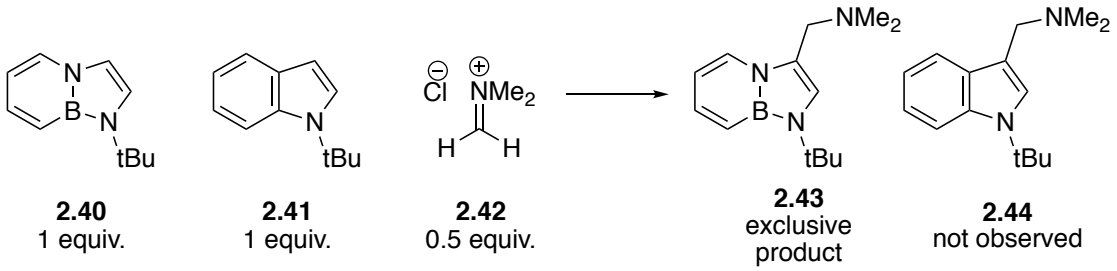
2.38 **2.39**

The incorporation of a BN-bond in place of a CC-double bond in indole results in an isoelectronic and isostructural analogue of indole (Figure 2.16). As described in the first chapter, while the BN-indole maintains the same steric profile as its carbonaceous counterpart, the electronic and chemical properties of the indole are altered.[52]

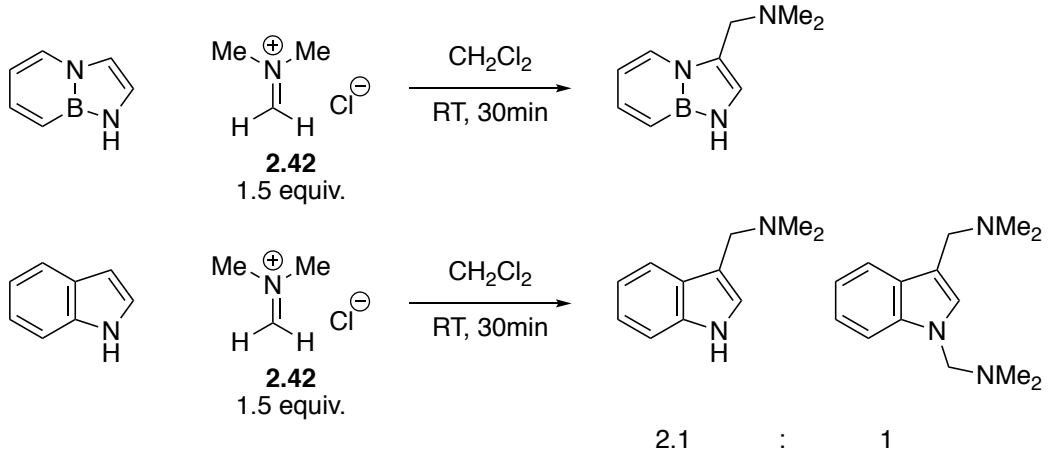
[52] Abbey, E. R.; Liu, S. Y. BN Isosteres of Indole. *Org. Biomol. Chem.* **2013**, *11*, 2060–2069.

Scheme 2.1 Mannich Reactions of Indole and BN-indole

a) Competition reaction between indole and BN-indole



b) Selectivity of the Mannich reaction of BN-indole and indole

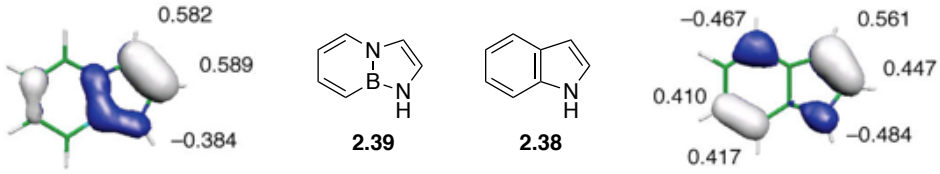


In 2010, the Liu group reported that the BN-indole demonstrates the same selectivity in electrophilic aromatic substitution reactions but is more nucleophilic.[53] In competition studies, it was shown that the BN-indole reacts preferentially over indole in a Mannich reaction with dimethyliminium chloride to form the gramine derivative (Scheme 2.1a). In a later study, it was also shown that the BN-indole did not undergo a second addition to the indole nitrogen, while in the indole case, a second addition does occur (Scheme 2.1b). [54] This report also showed that the BN-indole displays a higher HOMO-

[53] Abbey, E. R.; Zakharov, L. N.; Liu, S. Y. Electrophilic Aromatic Substitution of a BN Indole. *J. Am. Chem. Soc.* **2010**, *132*, 16340–16342.
[54] Chrostowska, A.; Xu, S.; Mazière, A.; Boknevitz, K.; Li, B.; Abbey, E. R.; Dargelos, A.; Graciaa, A.; Liu, S. Y. UV-Photoelectron Spectroscopy of BN Indoles: Experimental and Computational Electronic Structure Analysis. *J. Am. Chem. Soc.* **2014**, *136*, 11813–11820.

orbital coefficient at the C2 and C3 positions when compared with the indole, which corroborates the reactivity difference between the two indoles (Figure 2.17).

Figure 2.17 HOMO of Indoles with Selected π-orbital Coefficients



Chrostowska and Liu also reported a detailed study on the electronic structure of BN-indole through a combination of UV-photoelectron spectroscopy experiments and computational studies.[54] This study shows that the BN-indole has a smaller ground state dipole moment, is less aromatic (based on NICS values), and has a similar HOMO level energy compared to natural indole (Table 2.3).

Table 2.3 Selected Properties of Indole and BN-indole.

	pKa	λ_{abs}	λ_{emi}	Dipole	NICS (0)	NICS (1)	HOMO (eV)
Indole 2.38	20.95	268 nm	315 nm	2.177 D	–10.4, –12.9*	–11.7, –10.9*	7.9
BN-indole 2.39	30	293 nm	360 nm	1.512 D	–6.5, –11.0*	–7.8, –8.1*	8.05

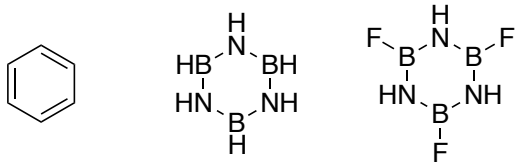
*five-membered ring

In 2011, the Liu group performed an NMR bracketing study to determine the p*K*a of the nitrogen-H.[55] They demonstrated that the p*K*a was significantly increased to ~30, whereas the natural indole has a p*K*a of 20.95. It was also shown that the BN-indole has red-shifted absorbance and emission spectra as well as a greater Stokes shift (Table 2.3). These changes in properties are considered a direct result of BN/CC isosterism.

[55] Abbey, E. R.; Zakharov, L. N.; Liu, S. Y. Boron in Disguise: The Parent Fused BN Indole. *J. Am. Chem. Soc.* **2011**, *133*, 11508–11511.

2.4.2. Computational Studies on Non-covalent Interactions with BN-heterocycles

Figure 2.18 Benzene and Borazines Explored in Computational Study



2.28 **2.38** **2.39**

Table 2.4 Interaction Energies (kcal/mol)

	Li^+	Na^+	K^+
2.38	–31.7	–19.2	–12.2
2.39	–20.3	–10.1	–4.9

While the vast majority of studies have focused on benzene and other more common heterocycles, very few studies have been performed on BN-containing heterocycles. In a study from 2005, the cation-π interactions of borazine **2.38** and its B-trifluoroborazine derivative **2.39** were explored computationally (Figure 2.18).[56] The results showed that the cation-π interaction with alkali cations was still favorable with borazine and its B-trifluoroborazine derivative (Table 2.4 vs. Table 2.1). They also found that the borazine compounds "puckered" in the presence of a cation with the three nitrogens being drawn closer and the borons pushed away. In the case of benzene, no such change in conformation is seen. They hypothesized the reasoning for this change is the attraction and repulsion between partial negative charge over the nitrogens and partial positive charge over the borons with the cation as well as the lower degree of aromaticity in borazine.

[56] Miao, R.; Yang, G.; Zhao, C.; Hong, J.; Zhu, L. Theoretical Study of Borazine and Its Fluoroderivatives: Aromaticity and Cation-π, Anion-π Interaction. *J. Mol. Struct. THEOCHEM* **2005**, *715*, 91–100.

Another computational work observed the same trend of borazine binding to various cations.[57] Within that study, it was found that the addition of electron-donating groups (EDG) to borazine increases the cation-π interaction and electron-withdrawing groups (EWG) decreases it, much like with benzene (Figure 2.19). Table 2.5 shows the calculated binding complexations of the explored arenes.

Figure 2.19 Benzene and Borazines Explored in Computational Study

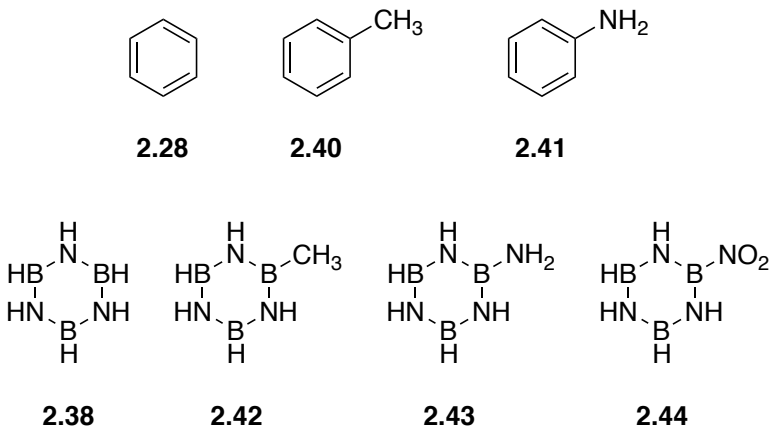


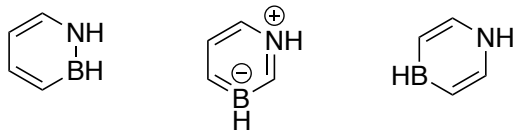
Table 2.5 Calculated Gas Phase Complexation Enthalpies (kcal/mol)

	Li^+	Na^+	K^+
2.28	–36.93	–24.12	–17.48
2.40	–44.22	–29.18	–23.33
2.41	–53.64	–38.76	–32.88
2.38	–31.77	–20.47	–15.03
2.42	–41.04	–27.25	–21.00
2.43	–54.19	–39.19	–32.06
2.44	8.30	10.66	11.63

[57] Bania, K. K.; Guha, A. K.; Bhattacharyya, P. K.; Sinha, S. Effect of Substituent and Solvent on Cation-π Interactions in Benzene and Borazine: A Computational Study. *Dalt. Trans.* **2014**, *43*, 1769–1784.

More recently, in 2018, the Korona group studied the interaction of Na^+ and Mg^{2+} with the three regioisomers of azaborine, 1,2-, 1,3-, and 1,4-azaborine, as well as mono-substituted variants through calculations (Figure 2.20).[58] It was found that the location of the cation bound to the ring was dependent on the location of the boron and nitrogen and the type of cation. In the case of benzene, the cation is always located above the center of the ring. This work, and another report from Korona,[59] suggest that more than the π-contribution needs to be considered, such as σ-polarization and induction energy, with regards to the azaborine compounds in order to accurately predict the interaction. In the case of benzene, usually the electrostatic π-contribution is sufficient.

Figure 2.20. Azaborine Isomers



1,2-azaborine 2.45 **1,3-azaborine 2.46** **1,4-azaborine 2.47**

2.5. Indole-scaffold Studies

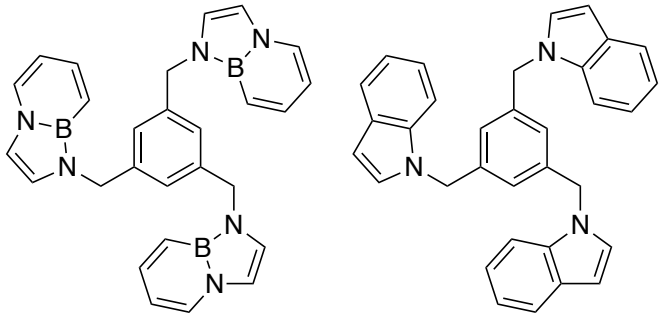
In Chapter 1, the incorporation of a BN-analogue of tryptophan into recombinant proteins was detailed. With this new ability to incorporate a BN-containing amino acid into proteins in our toolbox, learning about BN-indole's π-donating abilities could help us strategically apply our BN-tryptophan to other proteins. In order to probe the cation-π strength, we synthesized a BN-indole-containing scaffold **2.48** and determined the association constant by NMR, and compared the results with the natural indole scaffold **2.49** (Figure 2.21). This system was previously developed and explored by Hof and

[58] Yourdkhani, S.; Chojecki, M.; Korona, T. Interaction of Non-Polarizable Cations with Azaborine Isomers and Their Mono-Substituted Derivatives: Position, Induction, and Non-Classical Effects Matter. *ChemPhysChem* **2018**, *19*, 3092–3106.

[59] Yourdkhani, S.; Chojecki, M.; Korona, T. Substituent Effects in the So-Called Cation···π Interaction of Benzene and Its Boron-Nitrogen Doped Analogues: Overlooked Role of σ-Skeleton. *Phys. Chem. Chem. Phys.* **2019**, *21*, 6453–6466.

coworkers in order to study the cation-π binding ability of indole in aqueous and organic media.[60]

Figure 2.21 Scaffold Structures

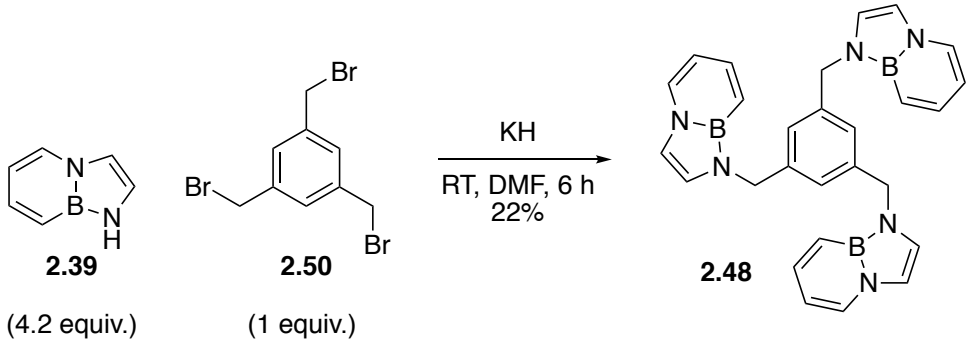


BN-indole scaffold 2.48 **CC-indole scaffold 2.49**

2.5.1. Synthesis of BN-indole Scaffold

In the Liu lab, we synthesized the CC-indole scaffold **2.49** as previously reported by Hof.[60] The synthesis of the BN-indole scaffold **2.48** was accomplished by adapting the method to synthesize the CC-indole scaffold **2.49** to the BN-system. Deprotonation of the BN-indole **2.39** with potassium hydride in dimethylformamide followed by addition of the 1,3,5-tribromomethylbenzene electrophile **2.50** led to the formation of the BN-scaffold **2.48** (Scheme 2.2).

Scheme 2.2 Synthesis of BN-scaffold **2.48**



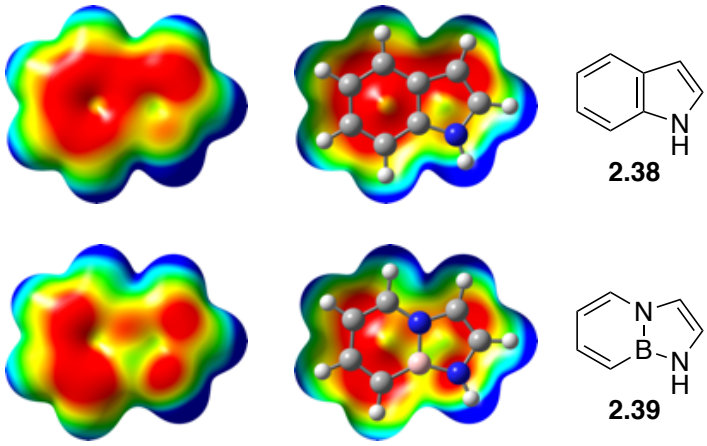
2.39 (4.2 equiv.) **2.50** (1 equiv.) **2.48**

[60] Whiting, A. L.; Neufeld, N. M.; Hof, F. A Tryptophan-Analog Host Whose Interactions with Ammonium Ions in Water Are Dominated by the Hydrophobic Effect. *Tetrahedron Lett.* **2009**, *50*, 7035–7037.

2.5.2. NMR Titrations

With the scaffolds in hand, the chemical shifts for scaffolds **2.48** and **2.49** were first unambiguously assigned via 2D NMR techniques (NOESY, HSQC, HMBC, COSY) prior to the titration experiments (see Sections 2.6.3 and 2.6.4). In order to compare the BN- and CC-scaffolds directly, we chose to look at the chemical shift changes associated with the C5 proton of the indole for the following reasons: 1) It is known that cations will bind preferentially to the six-membered ring of the indole,[25,61] thus a proton within the six-membered ring is appropriate, 2) the C5 proton signal is sharp and the chemical shift is easily assignable for both scaffolds,[62] and 3) the electrostatic potential (ESP) maps reveal a consistent negatively charged area located near the C5 and C6 positions for both scaffolds (Figure 2.22).

Figure 2.22 Electrostatic Potential Surfaces of CC- and BN-indole



[61] a) Brocchieri, L.; Karlin, S. Geometry of Interplanar Residue Contacts in Protein Structures. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 9297–9301. b) Aoki, K.; Murayama, K.; Nishiyama, H. Cation-π Interaction between the Trimethylammonium Moiety and the Aromatic Ring within Indole-3-Acetic Acid Choline Ester, a Model Compound for Molecular Recognition between Acetylcholine and Its Esterase: An X-Ray Study. *J. Chem. Soc. Chem. Commun.* **1995**, *764*, 2221–2222.

[62] The C5 and C6 proton chemicals shifts for the CC scaffold **2.49** overlap. However, the chemical shift changes upon binding to cations for the C5 and C6 protons are near identical, and the signal peak can be clearly assigned.

We them performed NMR titrations to determine the association constants for the binding of a lithium and sodium cation to the scaffolds (Figures 2.23 and 2.24). In both cases, a 0.9-1 mM solution of scaffold was made in acetonitrile-D3 and titrated with a 2.38-2.56 M solution of lithium iodide or a 771-820 mM solution of sodium iodide in acetonitrile-D3. The scaffold concentration was kept consistent between the scaffold solution and the salt solutions in order to eliminate any dilution effects that may arise. The resulting change in chemical shifts was fitted using the online resource provided by Thordarson,[63] and the association constants (K_a) were extracted.

Figure 2.23. Representative NMR spectra of LiI to BN-scaffold **2.48** titration

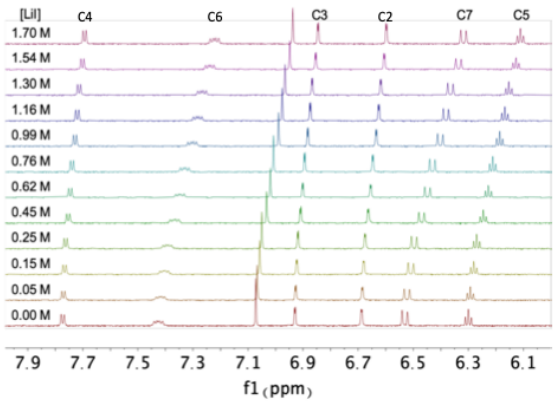
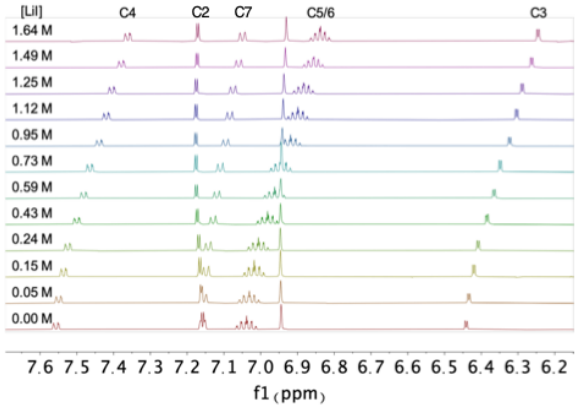


Figure 2.24. Representative NMR spectra of LiI to CC-scaffold **2.49** titration



We found BN-scaffold **2.48** gave a slightly lower K_a than the CC-scaffold **2.49** (Table 2.6). Binding with lithium iodide gave a slightly lower association constant for the BN-scaffold **2.48** with a K_a of 0.055 ± 0.003 M^{-1} compared to a K_a of 0.060 ± 0.004 M^{-1} for the CC-scaffold **2.49**. As expected from a solution phase cation-π binding system, the K_a for the binding of Na^+ (vs. Li^+) to BN-scaffold **2.48** was higher at 0.20 ± 0.01 M^{-1} and at 0.23 ± 0.04 M^{-1} for the CC-scaffold **2.49**. Overall, these results show that the

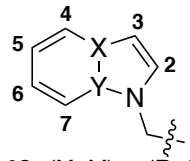
[63]a) See for binding model: Thordarson, P. Determining Association Constants from Titration Experiments in Supramolecular Chemistry. *Chem. Soc. Rev.* **2011**, *40*, 1305–1323. b) supramolecular.org. http://supramolecular.org (Accessed: 12/21/2019)

incorporation of the BN-bond into the indole motif results in a slightly lower cation-π binding ability.

Table 2.6 Association Constants of Lithium and Sodium Cation to Indole-Scaffolds

	Li^+	Na^+
BN-Scaffold **2.48**[1]	0.055 ± 0.003 M^{-1}	0.20 ± 0.01 M^{-1}
CC-Scaffold **2.49**[2]	0.060 ± 0.004 M^{-1}	0.23 ± 0.04 M^{-1}

[1] Measured with C5 ^{1}H NMR shifts
[2] Measured with C5/6 ^{1}H NMR shifts



2.48: (X–Y) = (B–N)
2.49: (X–Y) = (C=C)

2.5.3. Computational Studies

We also performed, in collaboration with Clovis Darrigan, a computational analysis of the gas-phase cation-π binding energy of Li^+, Na^+, and K^+ to the aromatic CC and BN-indole rings (Table 2.7). Three trends can be interpreted from the results. First, in all cases, the cation binds more strongly to the indole-π_6-system than the indole-π_5-system, and the interaction of both CC-indole-π_5 and BN-indole-π_5 systems with K^+ is unfavorable and unable to be measured. Second, the cations bind to CC-indole stronger than BN-indole by 2.5 kcal/mol for Li^+, 1.45 kcal/mol for Na^+, and 1.32 kcal/mol for K^+, with respect to indole-π_6, and 1.20 kcal/mol for Li^+ and 1.45 kcal/mol for Na^+, with respect to indole-π_5. Lastly, as is expected in the gas phase, the smaller Li^+ binds the strongest and the larger K^+ the weakest. The ESP maps (Figure 2.22) were also calculated and show that in both arenes there is a significant negative charge over the six-membered ring in both indoles, which is consistent with predicted stronger cation binding affinity for the π_6 ring system over the π_5 system for both CC- and BN-indole.

Table 2.7 Calculated Binding Energies of Indole and BN-indole with Alkali Cations

	$\Delta E_{Rig}(Li^+)$	$\Delta E_{Rig}(Na^+)$	$\Delta E_{Rig}(K^+)$
CC-Indole **2.38**-π_6	-48.3	-31.5	-23.0
BN-Indole **2.39**-π_6	-45.8	-30.1	-21.7
CC-Indole **2.38**-π_5	-42.9	-28.1	——
BN-Indole **2.39**-π_5	-41.7	-26.8	——

Energy values in kcal/mol. ΔE_{Rig} corresponds to the binding energy taking into account the basis set superposition error (BSSE).

2.5.4. Discussion

The NMR and computational results agree in some respects and disagree in others. In terms of the comparison between the BN- and CC-substitution, the smaller association constant seen for the BN-scaffold is corroborated by the calculated energy values. It is also important to note that the calculated energy values are gas phase values and therefore do not consider desolvation energies. Therefore, while the trend of the cation series in cation-π binding is opposite in the NMR versus the computational data, this is actually expected, as previously reported in Dougherty's early work.[42]

There also appears to be a correlation between the location of the negative charge in the ESP maps and the magnitude of the chemical shift change $\Delta\delta$ upon cation binding in the NMR titrations: for CC-indole **2.38**, the C2 position appears to be less electron rich according to the ESP map. This results in a small observed $\Delta\delta$ upon Li^+ or Na^+ binding (Figure 2.24)). For BN-indole **2.39**, the C2, C3, and C4 positions are relatively electron deficient, and consequently, the observed $\Delta\delta$ for these positions are smaller than the $\Delta\delta$ for the C5, C6, and C7 positions (Figure 2.23). It appears that the additional nitrogen atom in BN-indole **2.39** exerts its electronegative influence on the C3 and C4 positions giving rise

to a more non-uniform and position-dependent Δδ as a function of the titrated cation concentration relative to the CC indole **2.38**. Furthermore, we can see that the C7 position in BN indole **2.29** appears to be the most electron rich position, most likely due to the inductively donating adjacent boron, and as a consequence, the C7 proton experiences the largest Δδ upon titration with cation (Figure 2.23).

2.5.5. Conclusions

In conclusion, our results demonstrate that the incorporation of a B–N bond in indole has a slight effect on the arene's ability to bind simple cations as shown in our titration experiments. These results are corroborated through computations that indicate that there is slight difference in the cation-binding energy of the CC and BN-indoles. This knowledge could be used to help further predict the outcome of incorporating a BN-heterocycle in future chemical and biological applications.

2.6. Experimental Section

2.6.1. General Considerations

All oxygen- and moisture-sensitive manipulations were carried out under N_2 using either standard Schlenk techniques or a nitrogen-filled glovebox. DMF, Et_2O, CH_2Cl_2, and pentane were dried with a solvent purification system consisting of columns of molecular sieves under argon. Silica gel (230-400 mesh) was dried for 12 hours at 180 °C under high vacuum. Preparatory thin layer chromatography plates (TLC) were dried in an oven at 140 °C for 16 hours. Flash chromatography and TLC were performed with this silica gel under an inert atmosphere. All other chemicals and solvents were purchased and used as received.

NMR spectra were recorded on a Varian VNMRS 600 MHz or VNMRS 500 MHz spectrometer. Deuterated solvents were purchased from Cambridge Isotope Labs. ^{11}B NMR spectra were externally referenced to BF_3•Et_2O (δ 0). All NMR chemical shifts are reported in ppm relative to residual solvent for ^{1}H and ^{13}C NMR.

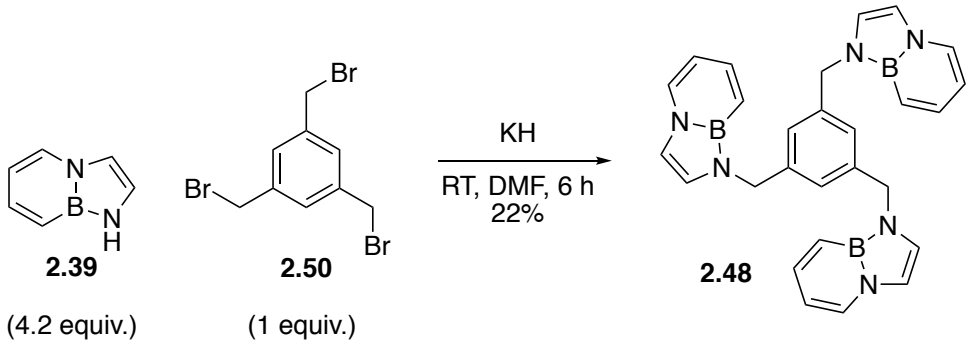
Infrared spectroscopy was performed on a Bruker ALPHA-Platinum FT-IR Spectrometer with ATR-sampling module.

High-resolution mass spectra were collected by Marek Domin on a JEOL AccuTOF instrument (JEOL USA, Peabody, MA), equipped with a DART ion source (IonSense, Inc., Danvers, MA) in positive ion mode at the Boston College Center for Mass Spectrometry.

2.6.2. Synthesis of BN-indole Scaffold

BN-indole **2.39** was prepared according to Abbey, et al.[55]

Synthesis of 2.48:

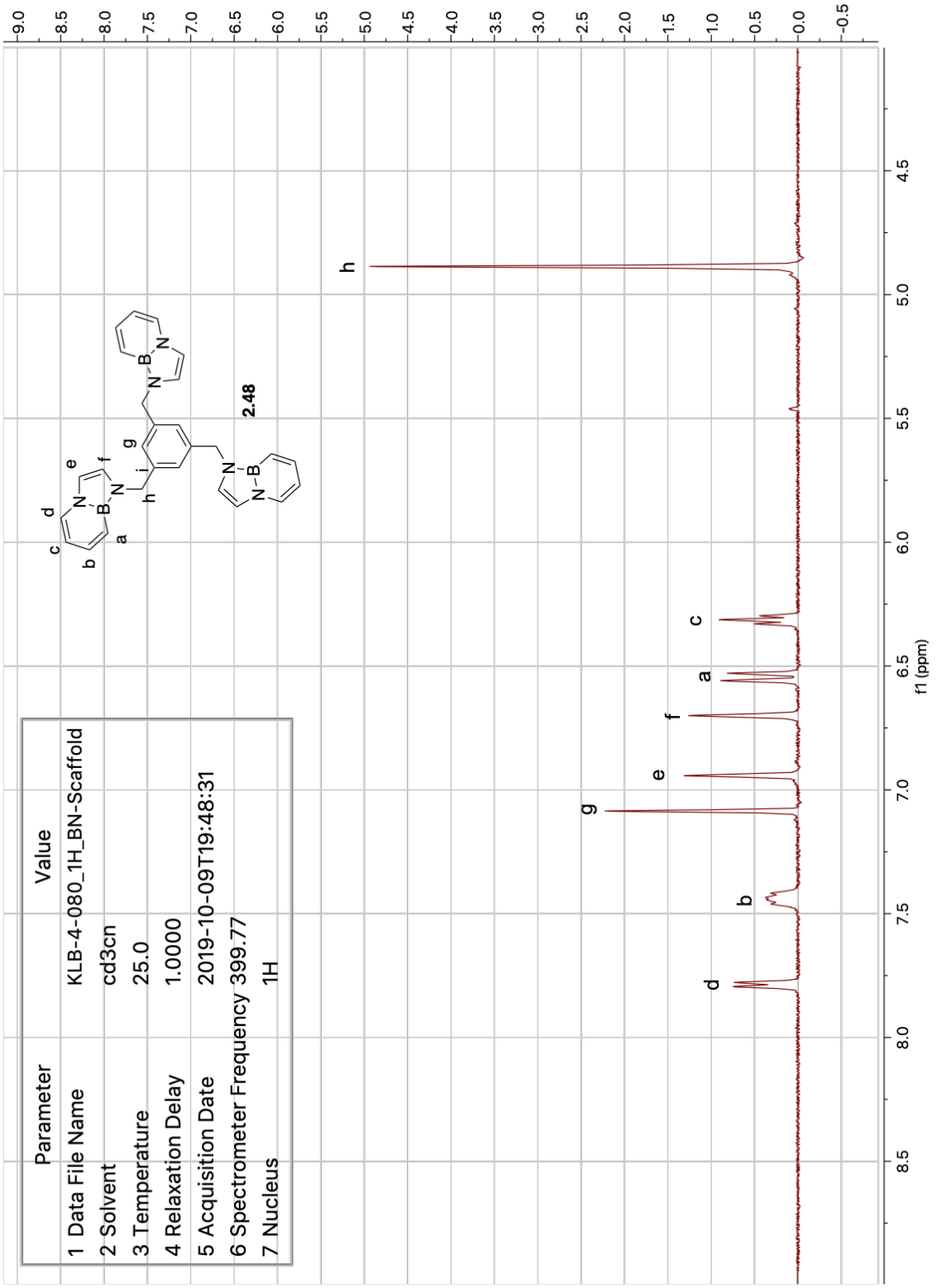


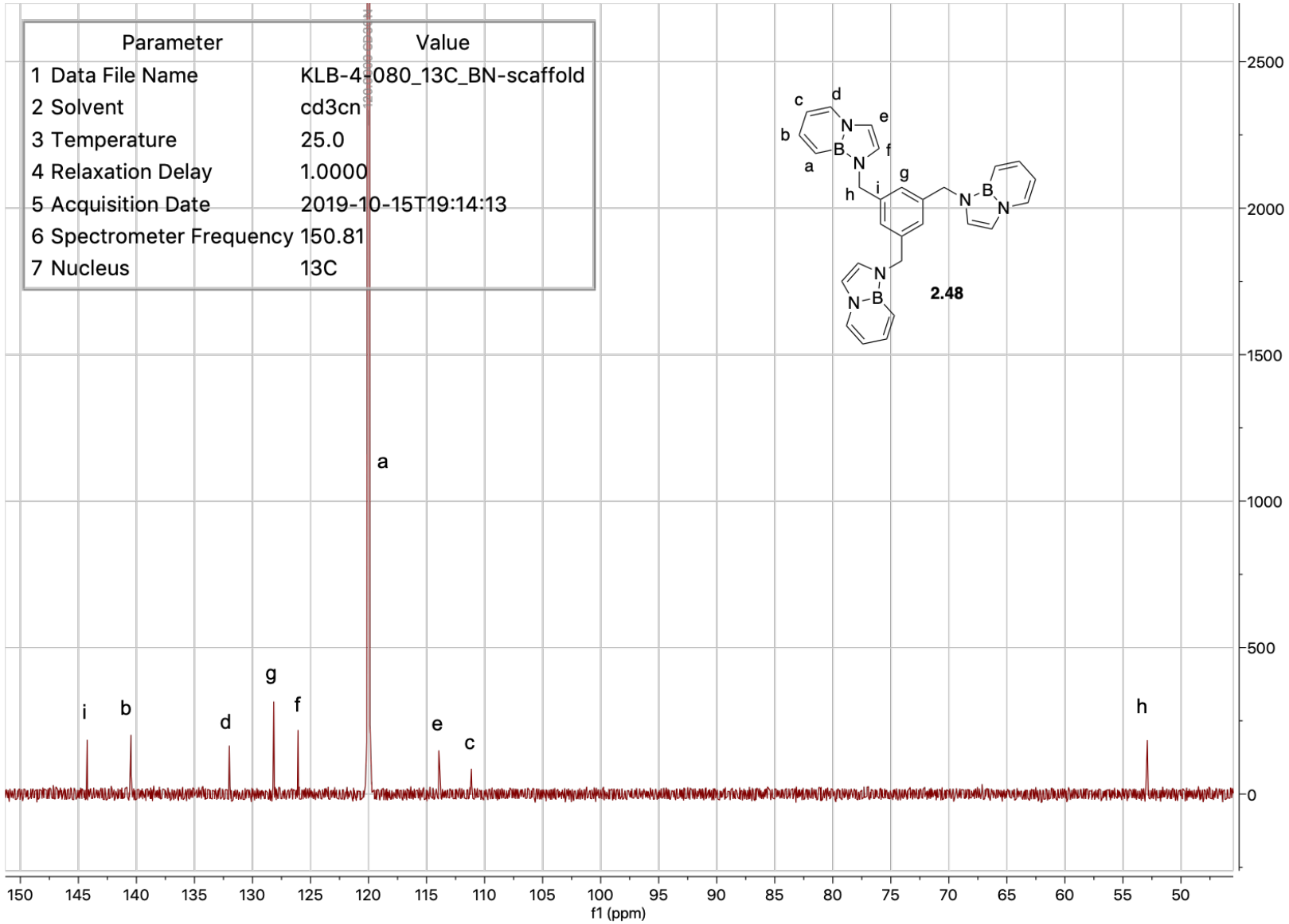
2.39 (4.2 equiv.) **2.50** (1 equiv.) **2.48**

In a glove box, an oven-dried 20 mL vial was charged with stir bar and BN-indole **2.39** (130 mg, 1.10 mmol, 4.2 equiv.) and dissolved in dimethylformamide (DMF) (6 mL). Potassium hydride (46.3 mg, 1.15 mmol, 4.4 equiv.) was added. The reaction mixture was allowed to stir at room temperature. After 2 hours, 1,3,5-tris(bromomethyl)benzene **2.50** (94 mg, 0.26 mmol, 1.0 equiv.) was dissolved in DMF (0.5 mL), added dropwise to the stirring reaction, and was allowed to continue stirring for 4 hours. The reaction was then quenched with water (15 mL) and transferred to a separatory funnel. Ethyl acetate (60 mL) was added and the layers separated. The organic layer was washed three times with water and once with brine, dried over $MgSO_4$, filtered, and volatiles removed under reduced pressure. The brown oil was then purified *via* preparatory thin layer chromatography in a glove box with 20% diethyl ether in pentane as eluent. The desired spot was scraped off with a spatula, the product was desorbed with dichloromethane (DCM), and concentrated in vacuo. Product was isolated as a tan solid (27 mg, 0.058 mmol, 22% yield). ^{1}H NMR (500 MHz, CD_2Cl_2) δ 7.73 (d, J = 6.5 Hz, 3H), 7.48 (dd, J = 11.6,

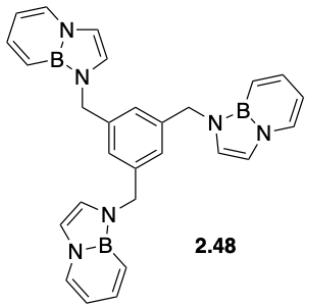
6.5 Hz, 3H), 7.02 (s, 3H), 6.89 (d, J = 2.2 Hz, 3H), 6.64 – 6.56 (m, 6H), 6.33 (t, J = 6.5 Hz, 3H), 4.89 (s, 6H). ^{13}C NMR (126 MHz, CD_2Cl_2) δ 141.65, 138.84, 129.52, 125.74, 123.64, 116.51, 112.12, 108.94, 51.01.^{11}B NMR (160 MHz, CD_2Cl_2) δ 25.3. IR (ATR) 2921.0, 2852.5, 1600.6, 1549.5, 1493.7, 1454.1, 1436.3, 1348.6, 1313.9, 1194.3, 1161.2, 1141.7, 1072.1, 1033.1, 983.3, 954.0, 865.9, 855.4, 815.8, 743.3, 726.6, 696.3, 674.4, 634.0, 613.7, 584.3, 535.6, 499.4, 485.6, 436.4 cm^{-1}. HRMS (DART+) calcd. for $C_{27}H_{28}B_3N_6$ $[M+H]^+$ 469.26491, found 469.26515.

2.6.3. NMR Assignment of BN-indole Scaffold

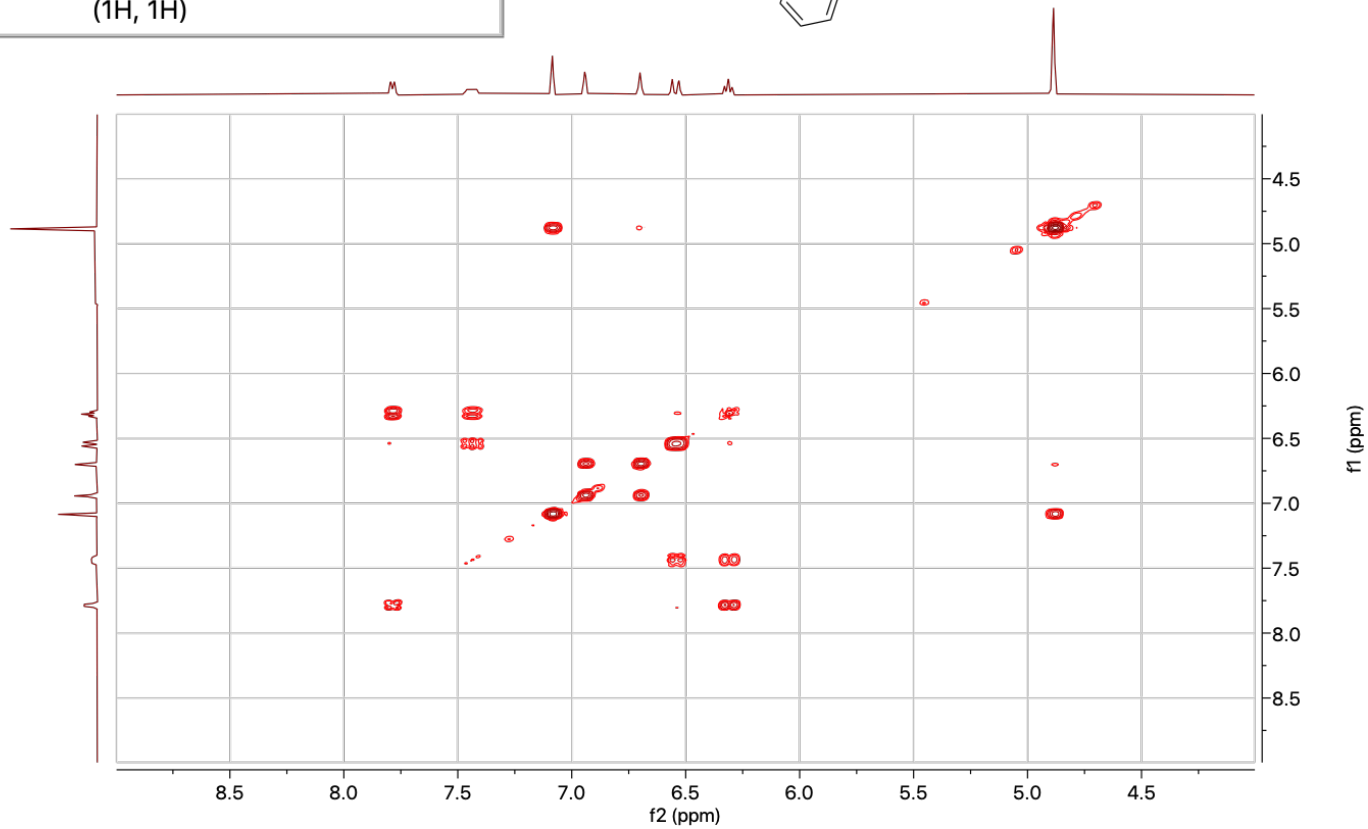



Parameter
Value
1 Data File Name
KLB-4-080_13C_BN-scaffold
2 Solvent
cd3cn
3 Temperature
25.0
4 Relaxation Delay
1.0000
5 Acquisition Date
2019-10-15T19:14:13
6 Spectrometer Frequency
150.81
7 Nucleus
13C
2.48
a
b
c
d
e
f
g
h
i
f1 (ppm)

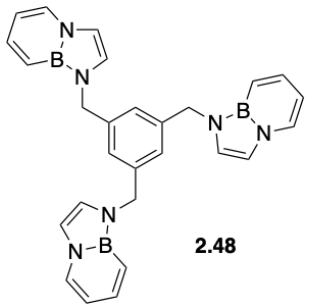
Parameter	Value
1 Data File Name	KLB-4-080_gCOSY_BN-Scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-09T21:14:26
6 Spectrometer Frequency	(399.77, 399.77)
7 Nucleus	(1H, 1H)



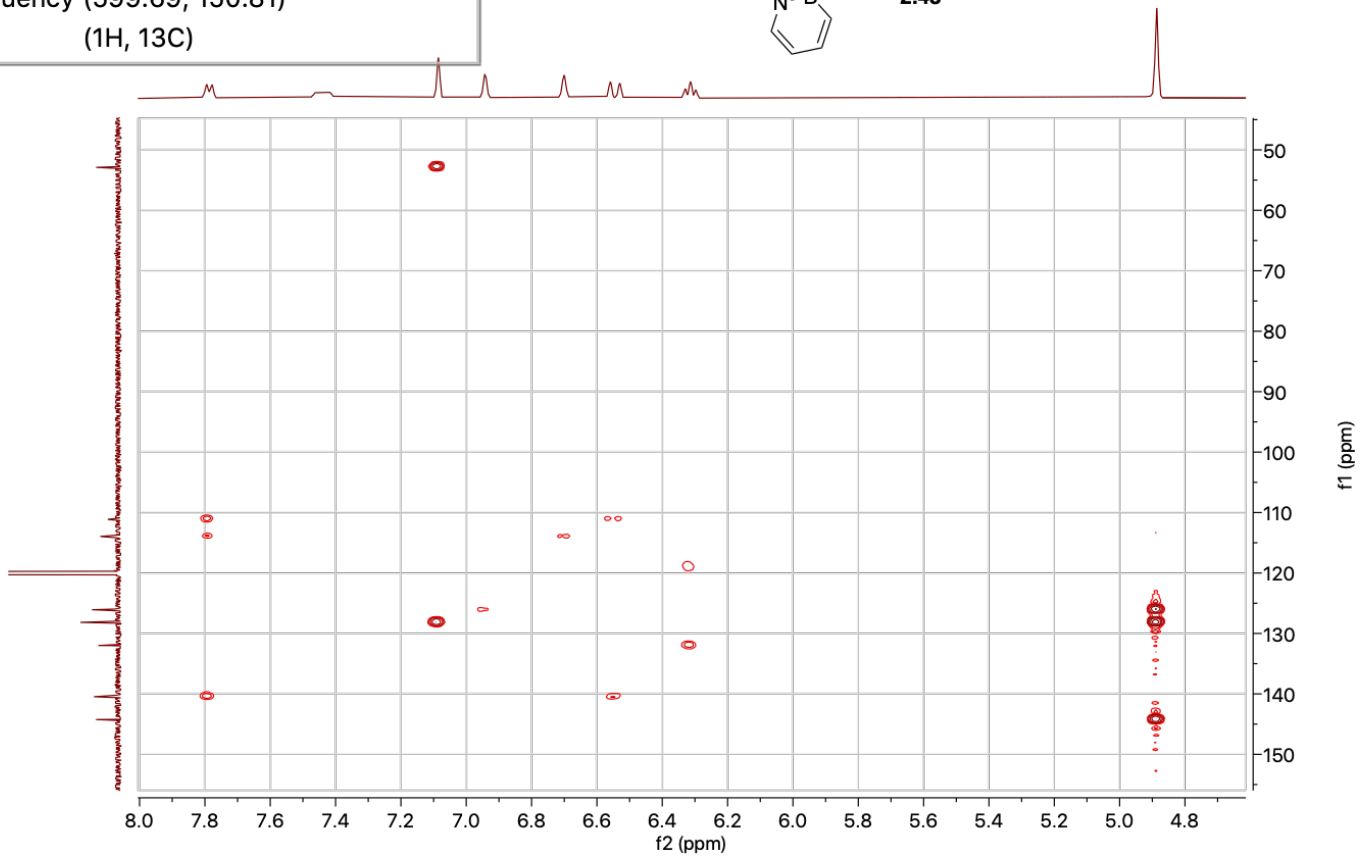
2.48



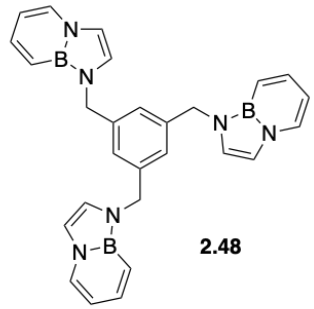
Parameter	Value
1 Data File Name	KLB-4-080_HMBC_BN-scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-16T10:34:10
6 Spectrometer Frequency	(599.69, 150.81)
7 Nucleus	(1H, 13C)



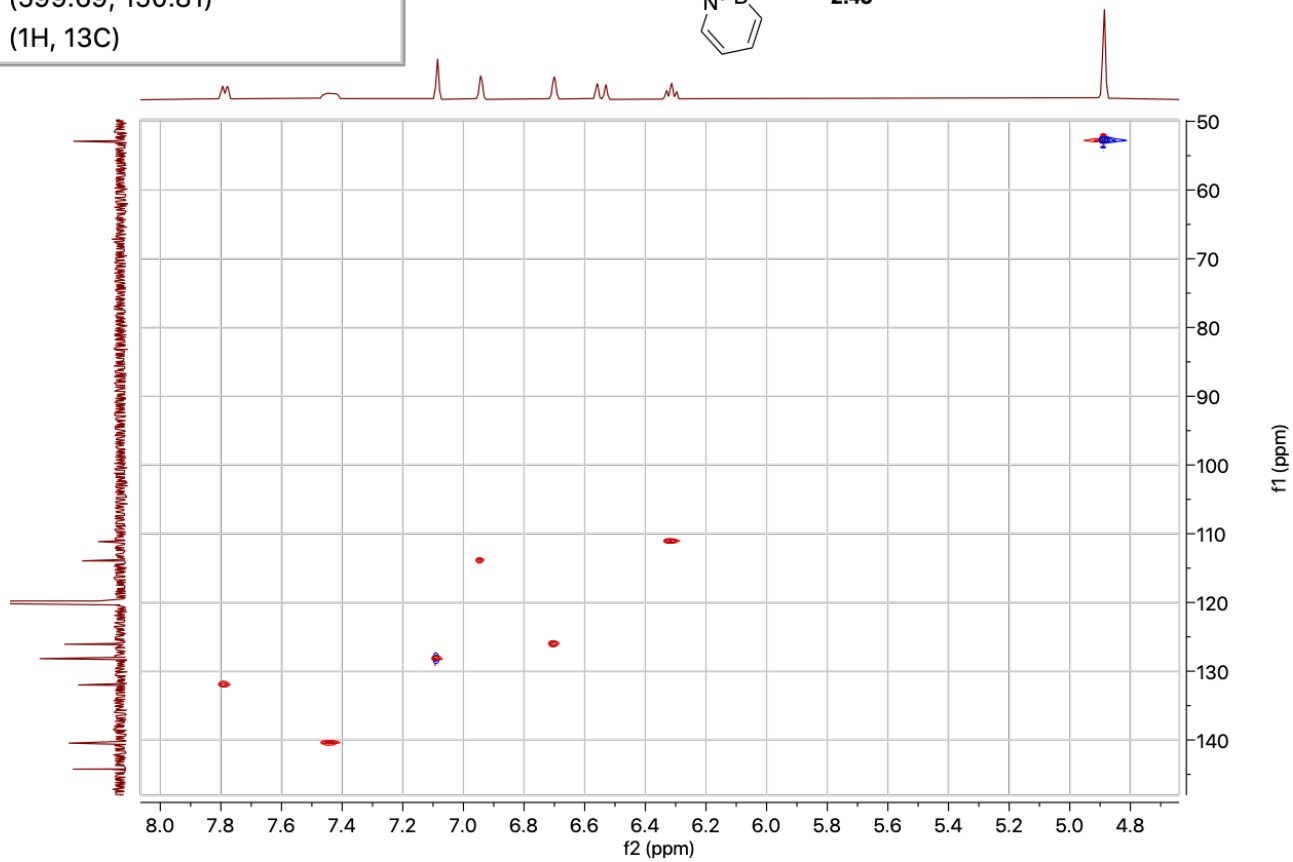
2.48



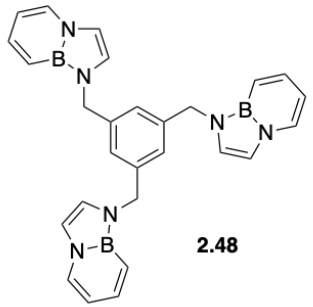
Parameter	Value
1 Data File Name	KLB-4-080_HSQC_BN-scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-15T19:46:47
6 Spectrometer Frequency	(599.69, 150.81)
7 Nucleus	(1H, 13C)



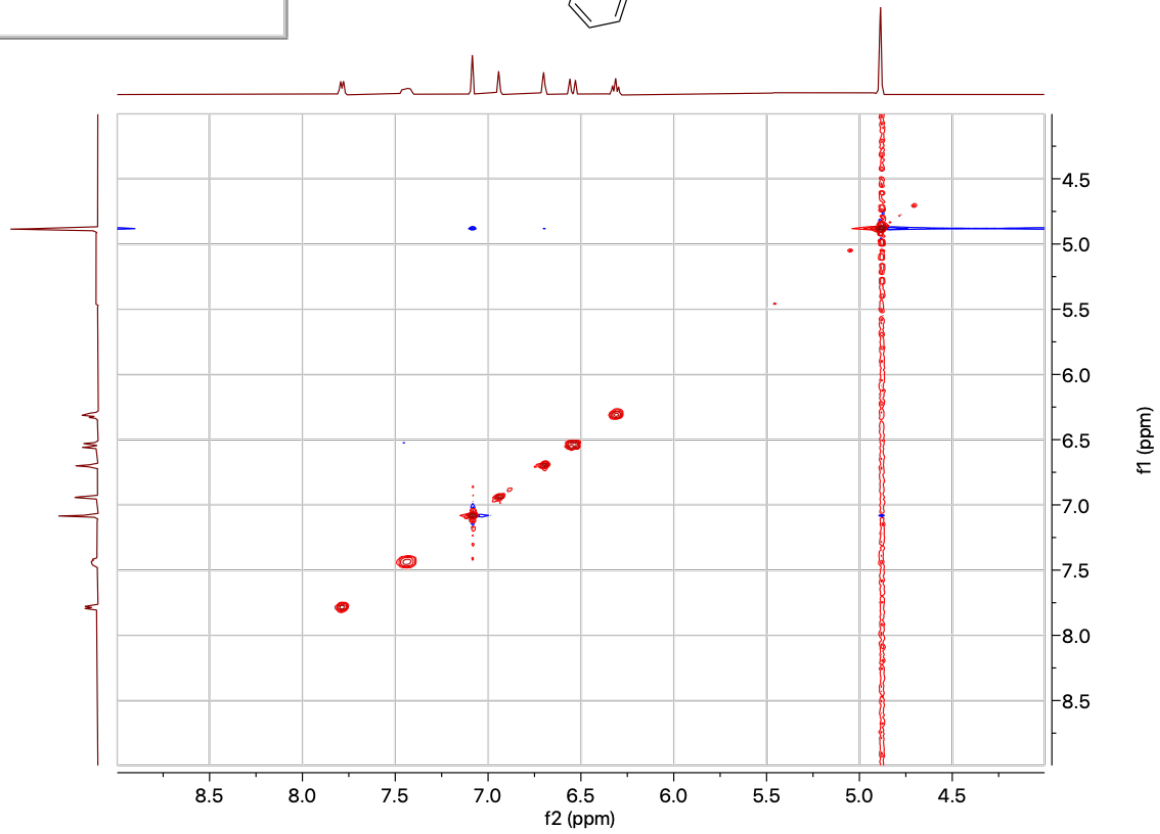
2.48



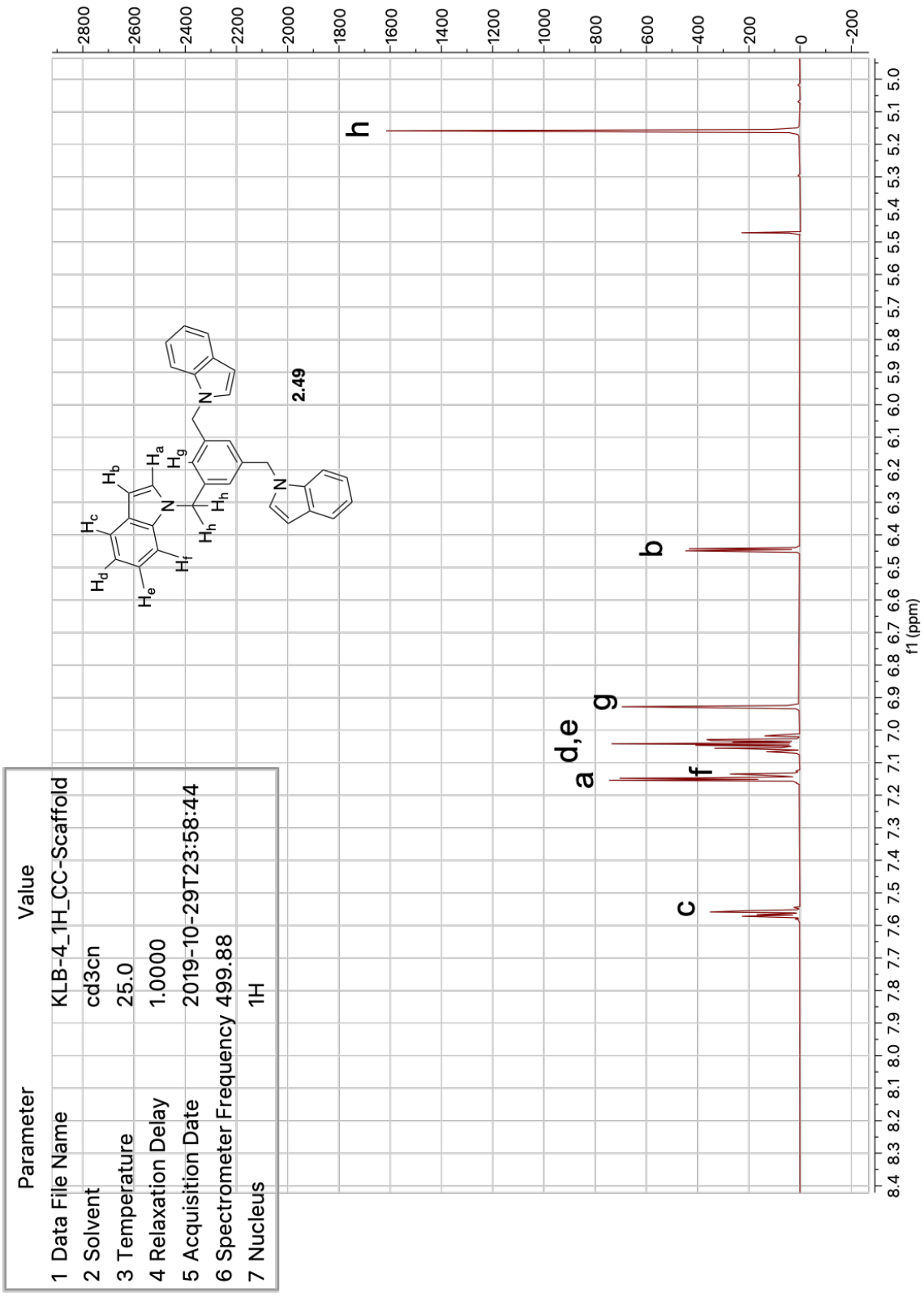
Parameter	Value
1 Data File Name	KLB-4-080_NOESY_BN-scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-09T19:56:24
6 Spectrometer Frequency	(399.77, 399.77)
7 Nucleus	(1H, 1H)

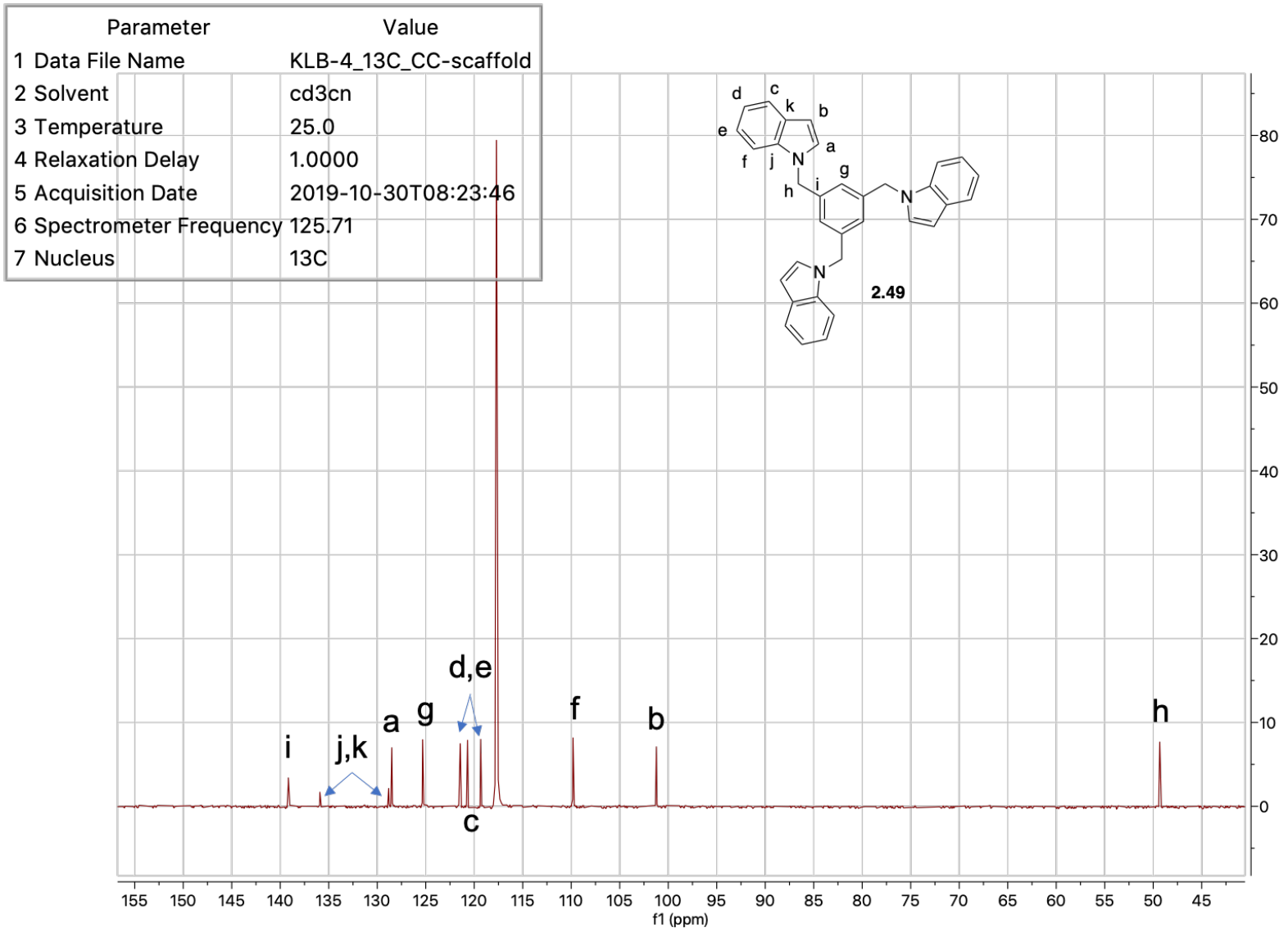


2.48

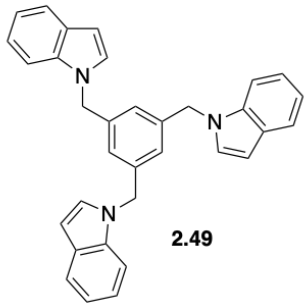


2.6.4. NMR Assignment of CC-indole Scaffold

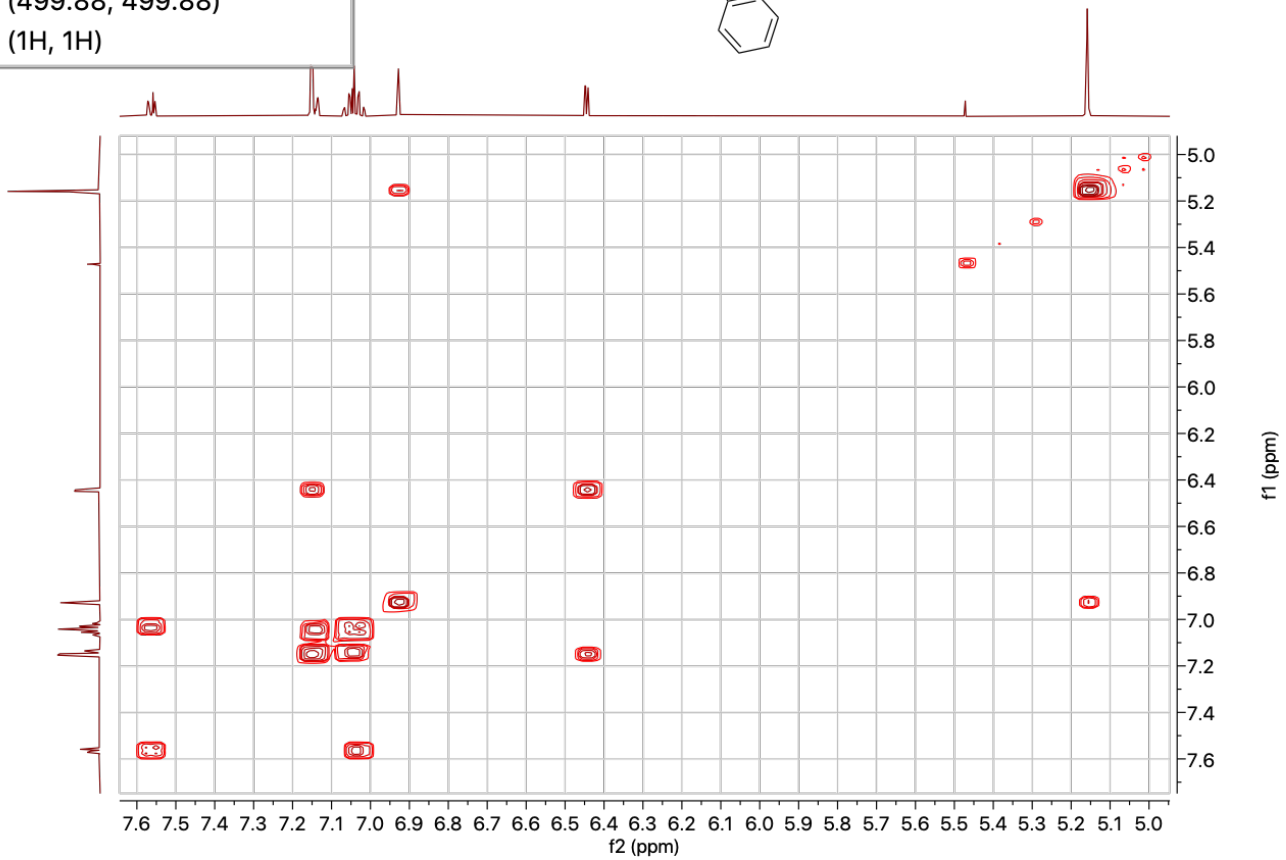



Parameter Value
1 Data File Name KLB-4_13C_CC-scaffold
2 Solvent cd3cn
3 Temperature 25.0
4 Relaxation Delay 1.0000
5 Acquisition Date 2019-10-30T08:23:46
6 Spectrometer Frequency 125.71
7 Nucleus 13C
2.49
i
j,k
a
g
d,e
c
f
b
h
f1 (ppm)

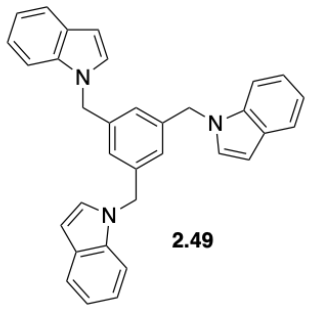
Parameter	Value
1 Data File Name	KLB-4_gCOSY_CC-scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-30T09:04:41
6 Spectrometer Frequency	(499.88, 499.88)
7 Nucleus	(1H, 1H)



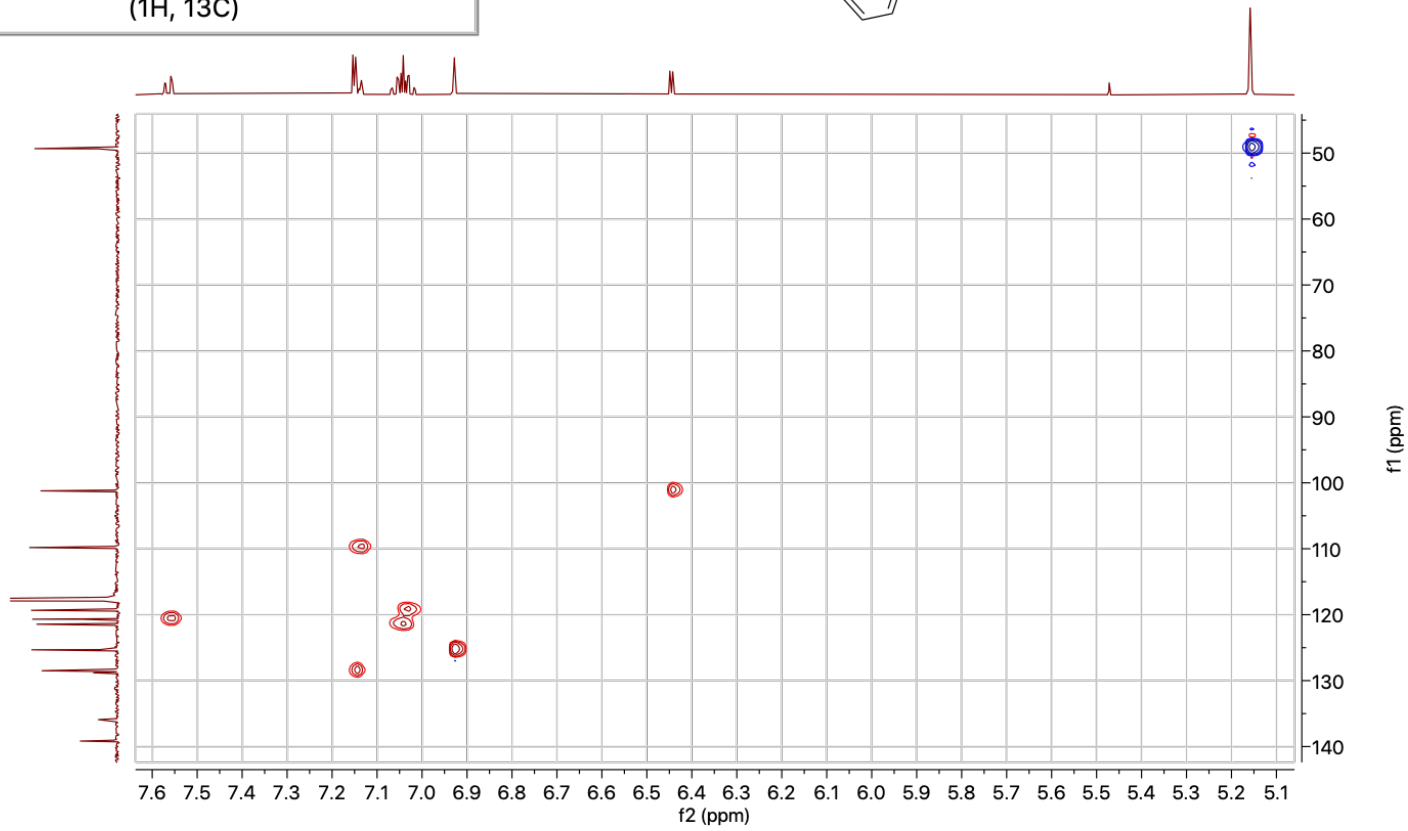
2.49



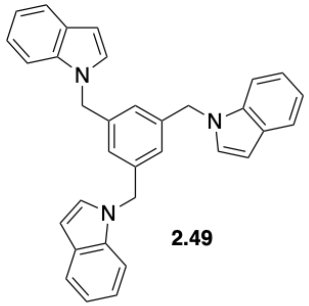
Parameter	Value
1 Data File Name	KLB-4_gHSQC_CC-scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-30T03:05:09
6 Spectrometer Frequency	(499.88, 125.71)
7 Nucleus	(1H, 13C)



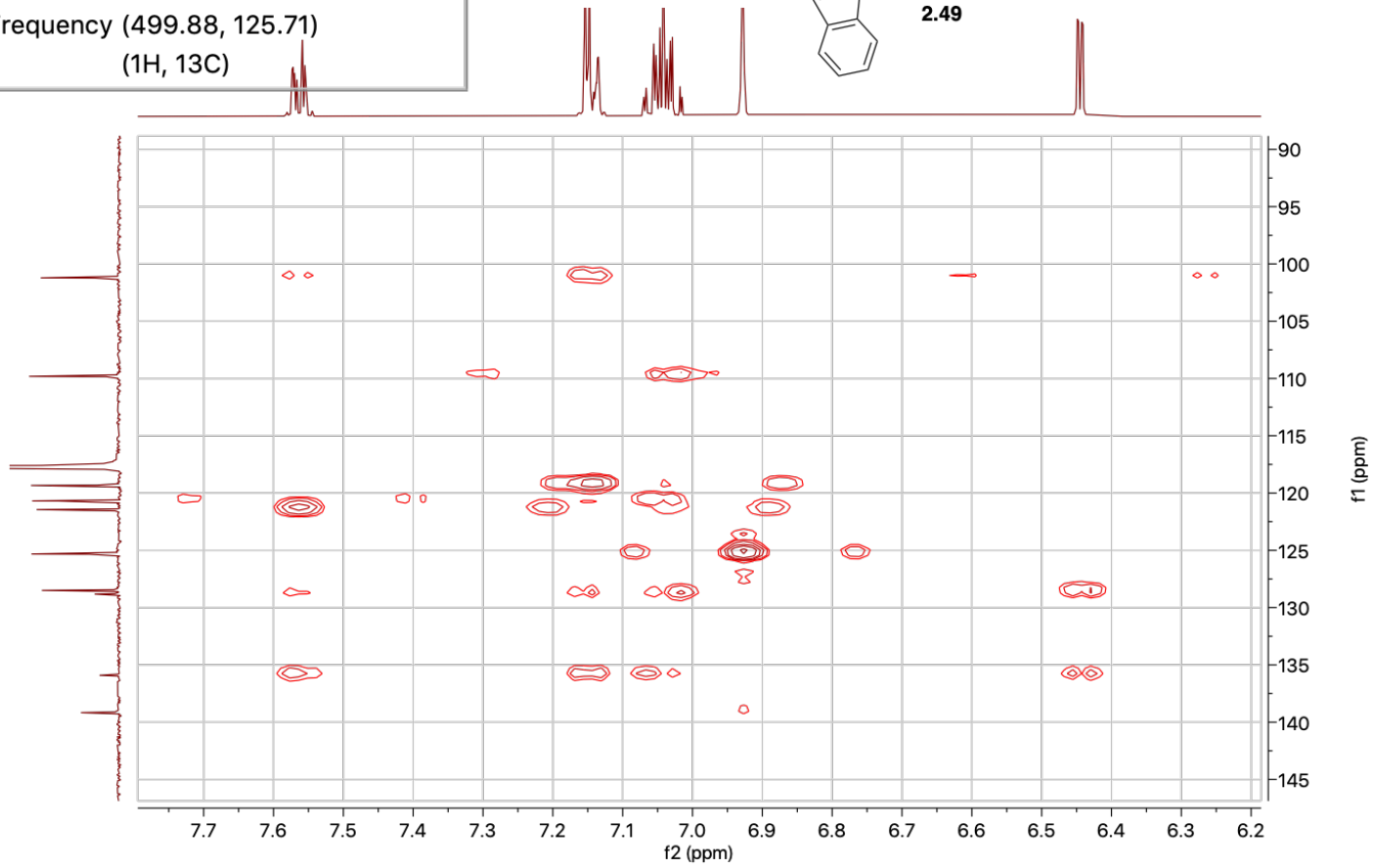
2.49

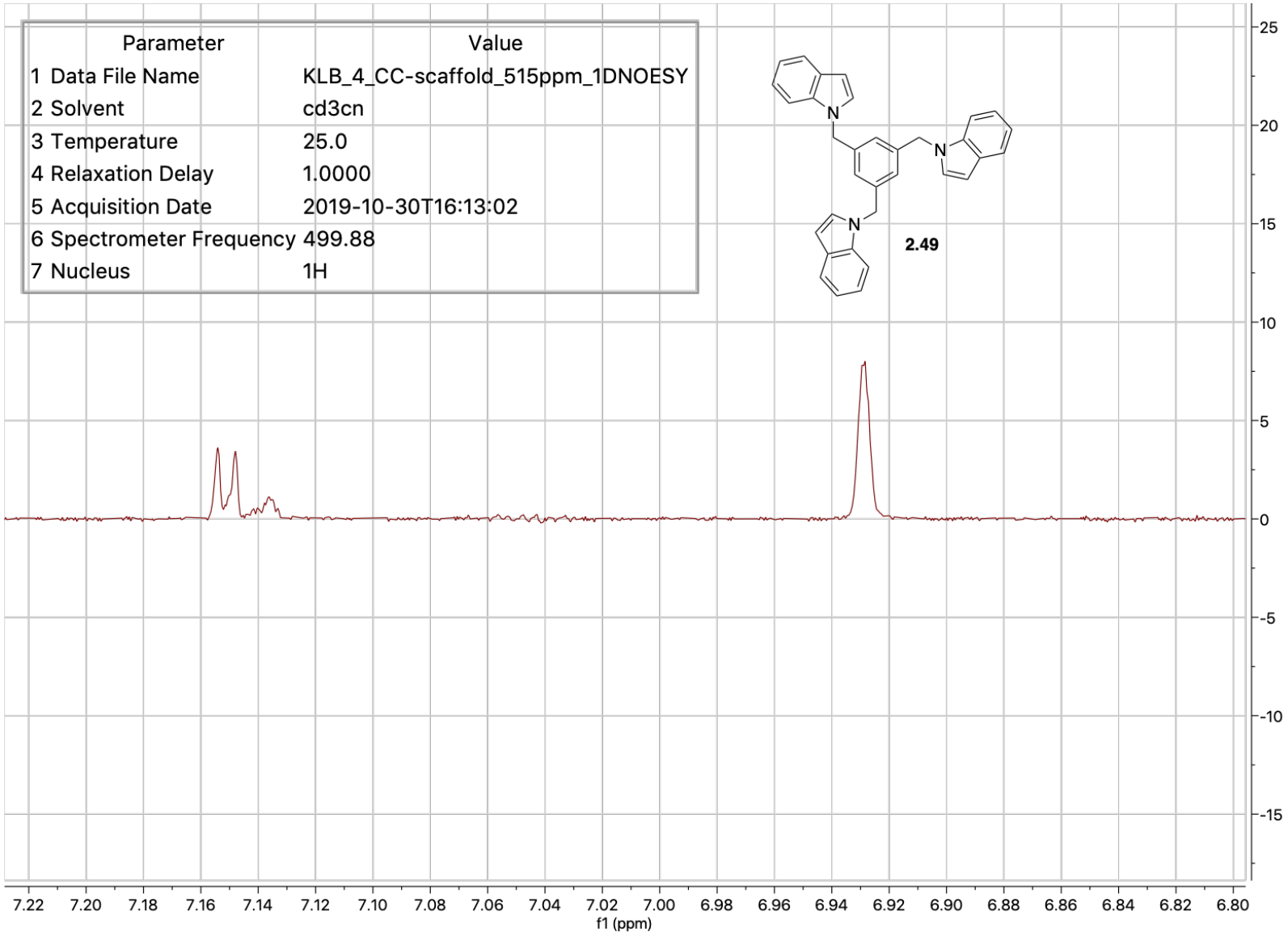


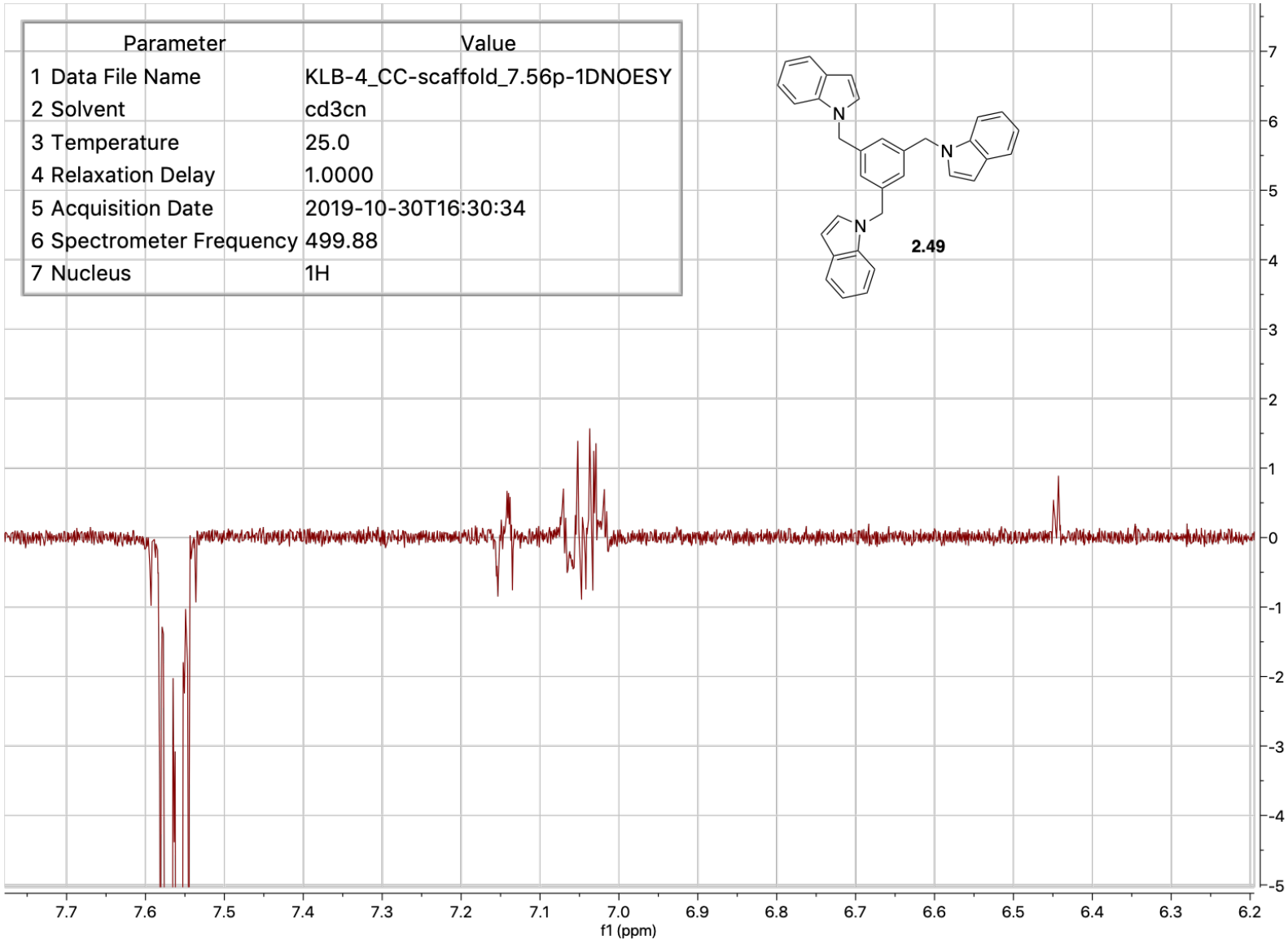
Parameter	Value
1 Data File Name	KLB-4_gHMBC_CC-scaffold
2 Solvent	cd3cn
3 Temperature	25.0
4 Relaxation Delay	1.0000
5 Acquisition Date	2019-10-30T08:02:14
6 Spectrometer Frequency	(499.88, 125.71)
7 Nucleus	(1H, 13C)



2.49

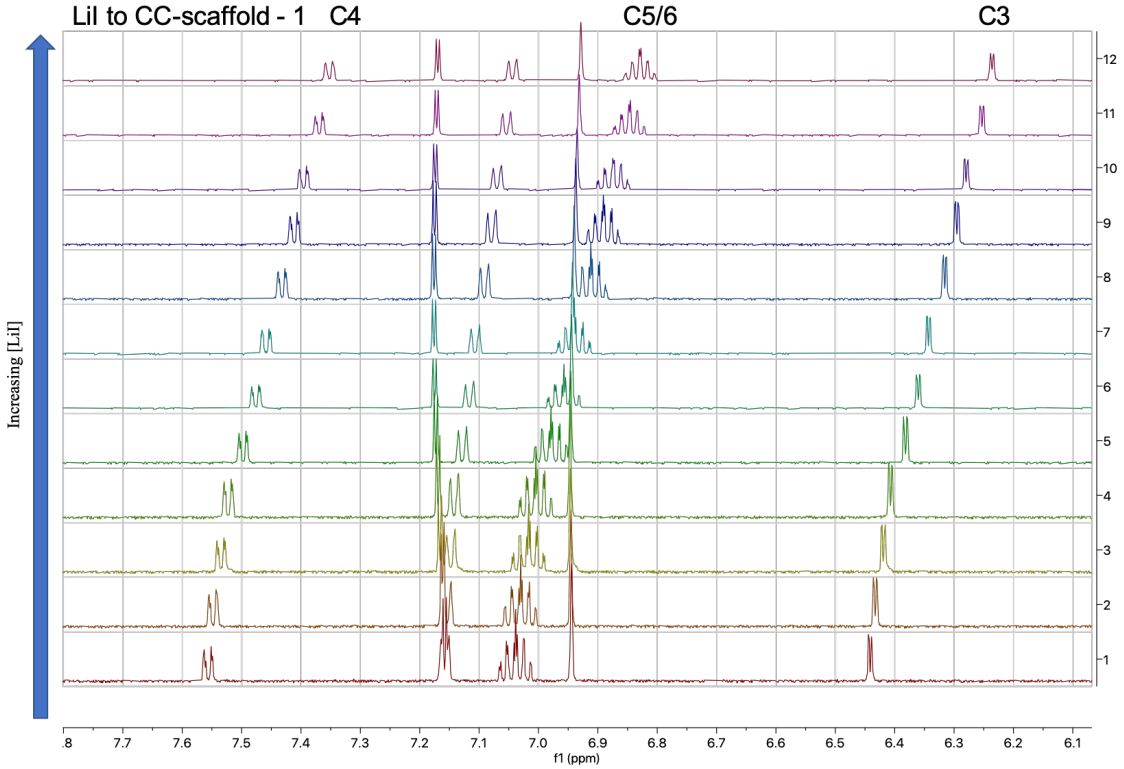



Parameter Value
1 Data File Name KLB_4_CC-scaffold_515ppm_1DNOESY
2 Solvent cd3cn
3 Temperature 25.0
4 Relaxation Delay 1.0000
5 Acquisition Date 2019-10-30T16:13:02
6 Spectrometer Frequency 499.88
7 Nucleus 1H
N
N
N
2.49
25
20
15
10
5
0
-5
-10
-15
7.22 7.20 7.18 7.16 7.14 7.12 7.10 7.08 7.06 7.04 7.02 7.00 6.98 6.96 6.94 6.92 6.90 6.88 6.86 6.84 6.82 6.80
f1 (ppm)


Parameter Value
1 Data File Name KLB-4_CC-scaffold_7.56p-1DNOESY
2 Solvent cd3cn
3 Temperature 25.0
4 Relaxation Delay 1.0000
5 Acquisition Date 2019-10-30T16:30:34
6 Spectrometer Frequency 499.88
7 Nucleus 1H
N
N
N
2.49
7.7
7.6
7.5
7.4
7.3
7.2
7.1
7.0
6.9
6.8
6.7
6.6
6.5
6.4
6.3
6.2
f1 (ppm)
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5

2.6.5. NMR Titration Spectra and Peak Data

LiI to CC-Scaffold 2.49: Titration 1



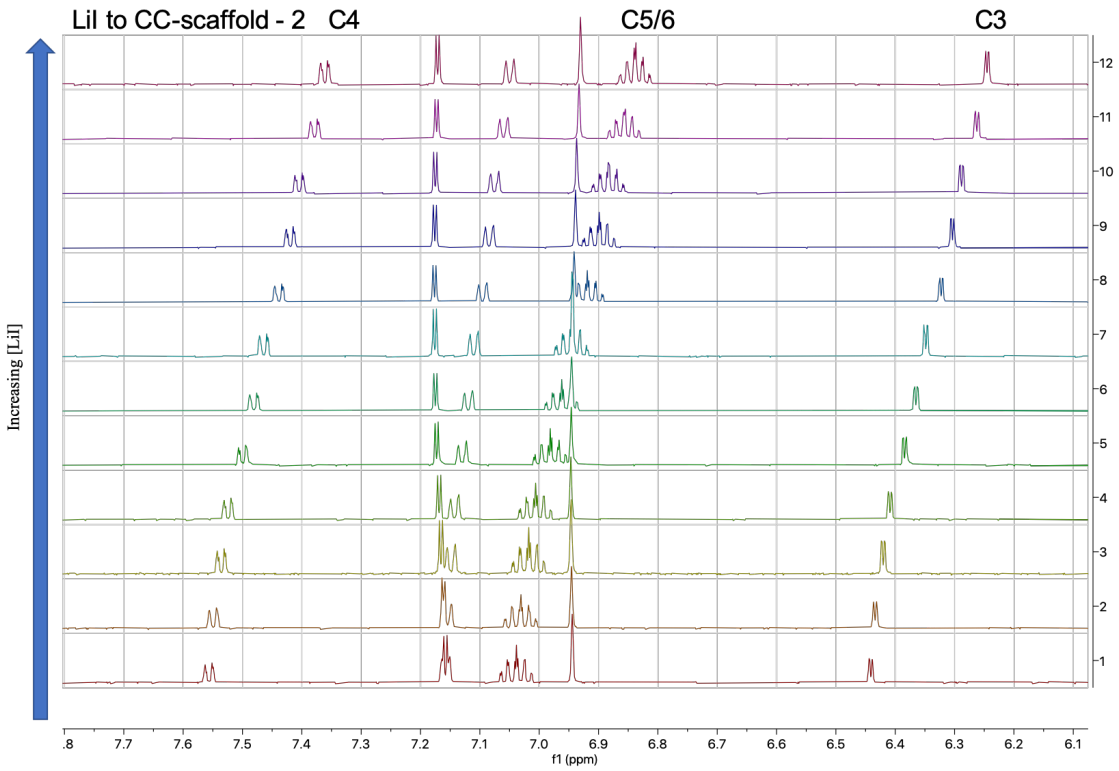
[LiI] Stock (M) **[CC-scaffold] (M)**

2.535 0.00103

[LiI] (M)	C4 (δ)	|C4 ($\Delta\delta$)|	C5/C6 (δ)	|C5/C6 ($\Delta\delta$)|	C3 (δ)	|C3 ($\Delta\delta$)|
0.000	7.556142	0.000000	6.943974	0.000000	6.440985	0.000000
0.055	7.547886	0.008256	6.934208	0.009766	6.432748	0.008237
0.158	7.534667	0.021475	6.923134	0.020840	6.419299	0.021686
0.253	7.522397	0.033745	6.910922	0.033052	6.406954	0.034031
0.461	7.497326	0.058815	6.896899	0.047075	6.381709	0.059276
0.634	7.476051	0.080090	6.882412	0.061562	6.360276	0.080709
0.780	7.458818	0.097323			6.343124	0.097861
1.014	7.431852	0.124289	6.866386	0.077588	6.316016	0.124969
1.193	7.411545	0.144597	6.853966	0.090008	6.295491	0.145494
1.334	7.395900	0.160242	6.840152	0.103822	6.279426	0.161559
1.584	7.369182	0.186960	6.819534	0.124440	6.253326	0.187659
1.748	7.351856	0.204286	6.801469	0.142505	6.236054	0.204931
2.535	7.274398	0.281743	6.746772	0.197202	6.159424	0.281561

All shifts in ppm

LiI to CC-Scaffold 2.49: Titration 2



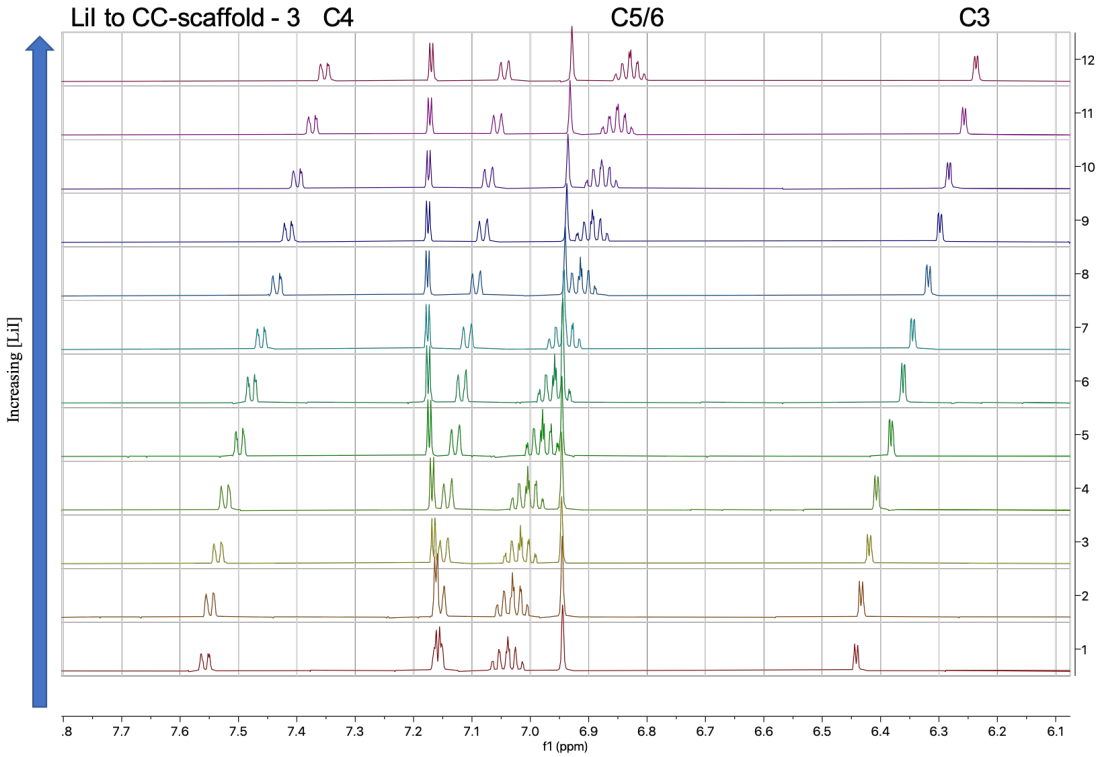
[LiI] Stock (M) **[CC-scaffold] (M)**

2.376 0.00103

[LiI] (M)	C4 (δ)	|C4 ($\Delta\delta$)|	C5/C6 (δ)	|C5/C6 ($\Delta\delta$)|	C3 (δ)	|C3 ($\Delta\delta$)|
0.000	7.556464	0.000000	7.038319	0.000000	6.441086	0.000000
0.052	7.548728	0.007737	7.028194	0.010125	6.433422	0.007663
0.148	7.535648	0.020816	7.017718	0.020601	6.420261	0.020825
0.238	7.524238	0.032227	7.006057	0.032262	6.408660	0.032425
0.432	7.499775	0.056690	6.981338	0.056981	6.384192	0.056894
0.594	7.480709	0.075756	6.961984	0.076335	6.365077	0.076009
0.731	7.464211	0.092254			6.348458	0.092628
0.950	7.438634	0.117830	6.918862	0.119457	6.322794	0.118292
1.118	7.419255	0.137210	6.898980	0.139339	6.303443	0.137643
1.250	7.404322	0.152142	6.883628	0.154691	6.288512	0.152574
1.485	7.378322	0.178142	6.856837	0.181482	6.262724	0.178362
1.638	7.361243	0.195221	6.838594	0.199726	6.245561	0.195525

All shifts in ppm

LiI to CC-Scaffold 2.49: Titration 3

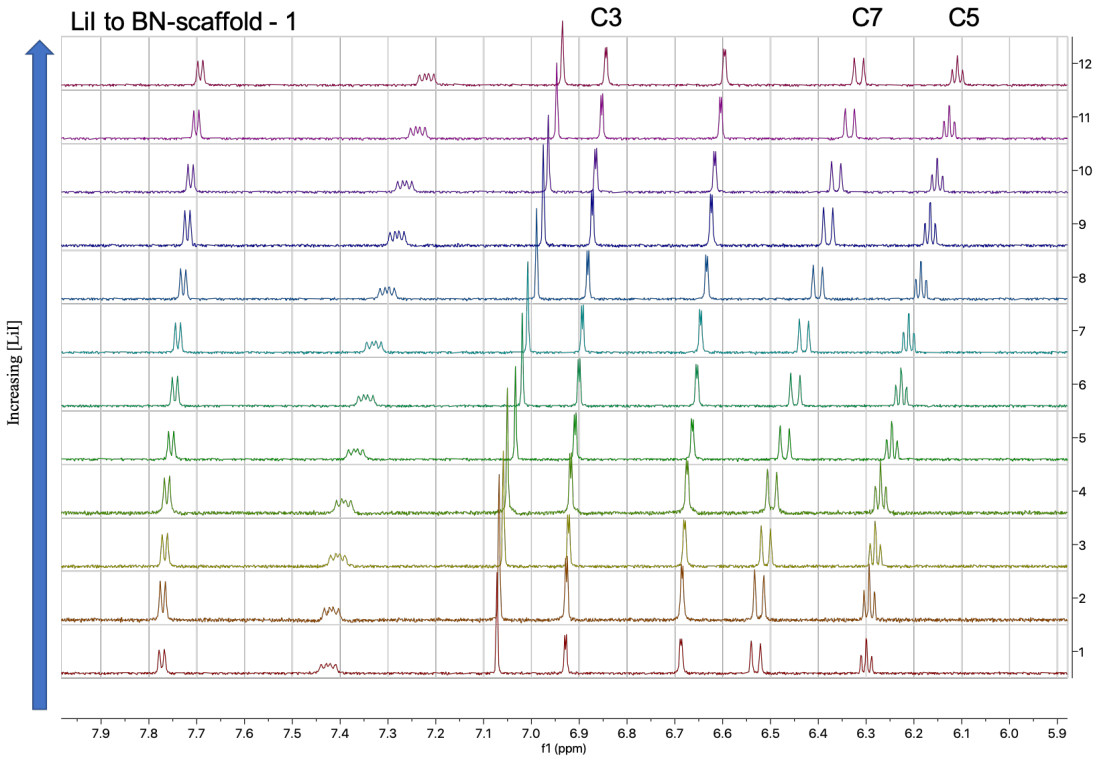


[LiI] Stock (M)	[CC-scaffold] (M)
2.442	0.00095

[LiI] (M)	C4 (δ)	|C4 ($\Delta\delta$)|	C5/C6 (δ)	|C5/C6 ($\Delta\delta$)|	C3 (δ)	|C3 ($\Delta\delta$)|
0.000	7.556931	0.000000	7.038703	0.000000	6.441480	0.000000
0.053	7.548255	0.008676	7.030824	0.007880	6.433359	0.008121
0.153	7.534815	0.022116	7.016986	0.021718	6.419298	0.022183
0.244	7.522400	0.034531	7.004672	0.034032	6.407261	0.034219
0.444	7.497547	0.059384	6.979218	0.059486	6.381963	0.059517
0.611	7.477432	0.079499	6.958611	0.080093	6.361669	0.079811
0.751	7.460906	0.096025			6.345062	0.096418
0.977	7.434208	0.122723	6.914436	0.124268	6.318441	0.123039
1.149	7.414483	0.142448	6.893834	0.144870	6.298421	0.143059
1.285	7.398725	0.158206	6.878109	0.160594	6.283099	0.158381
1.526	7.373027	0.183904	6.850923	0.187781	6.257157	0.184324
1.684	7.352312	0.204619	6.829298	0.209405	6.236508	0.204972

All shifts in ppm

LiI to BN-Scaffold 2.48: Titration 1



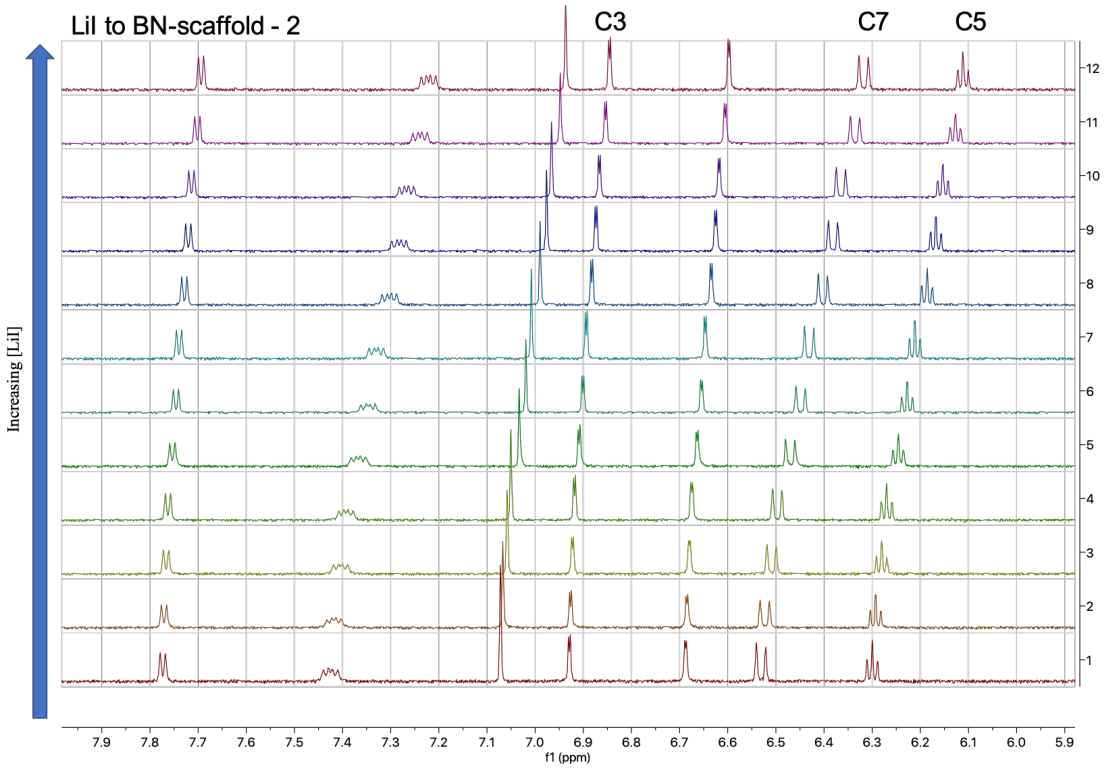
[LiI] Stock (M) **[BN-scaffold] (M)**

2.549 0.00098

[LiI] (M)	C3 (δ)	|C3 ($\Delta\delta$)|	C7 (δ)	|C7 ($\Delta\delta$)|	C5 (δ)	|C5 ($\Delta\delta$)|
0.000	6.928642	0.000000	6.530723	0.000000	6.299486	0.000000
0.055	6.926693	0.001949	6.523508	0.007216	6.293090	0.006396
0.159	6.922358	0.006284	6.509570	0.021153	6.280696	0.018790
0.255	6.917896	0.010745	6.496644	0.034079	6.269310	0.030176
0.463	6.908647	0.019994	6.470143	0.060580	6.246143	0.053344
0.637	6.900363	0.028279	6.447980	0.082743	6.226615	0.072871
0.784	6.893497	0.035144	6.429991	0.100732	6.210758	0.088728
1.019	6.881950	0.046692	6.401046	0.129677	6.185245	0.114241
1.199	6.872681	0.055961	6.379191	0.151532	6.166140	0.133347
1.341	6.865398	0.063243	6.362450	0.168274	6.151403	0.148084
1.593	6.852864	0.075778	6.334027	0.196697	6.126004	0.173483
1.758	6.843961	0.084681	6.314939	0.215785	6.109045	0.190441

All shifts in ppm

LiI to BN-Scaffold 2.48: Titration 2



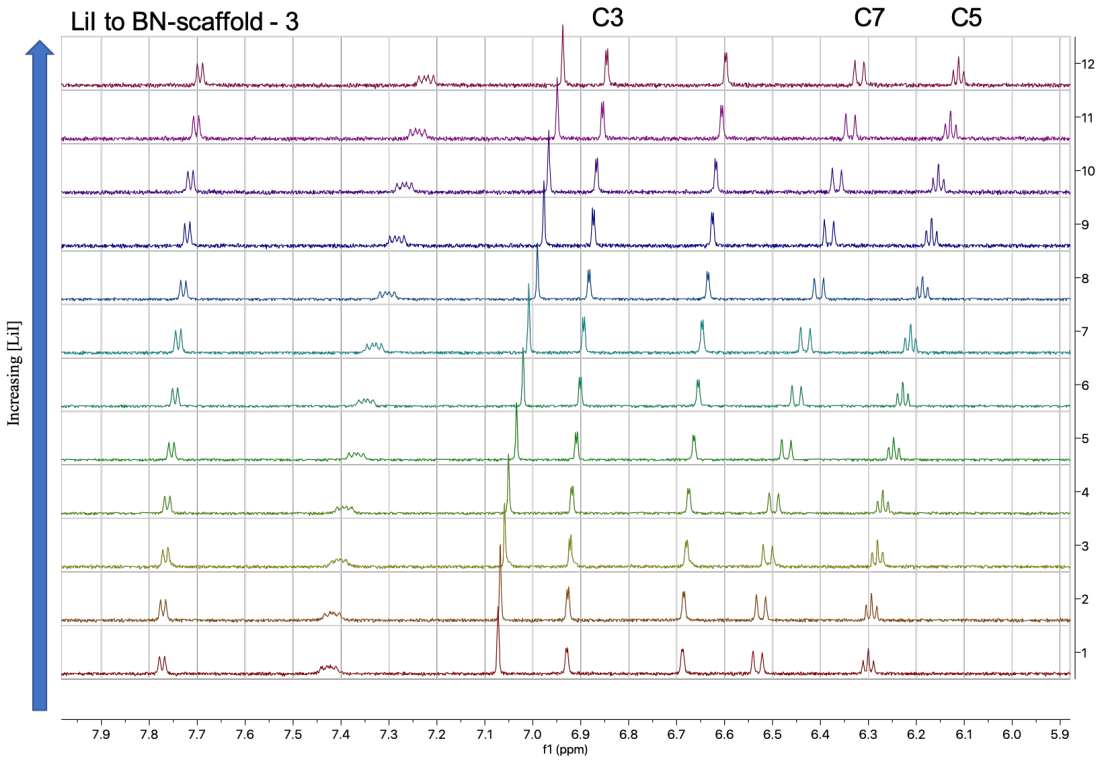
[LiI] Stock (M) **[BN-scaffold] (M)**

2.464 0.00098

[LiI] (M)	C3 (δ)	|C3 ($\Delta\delta$)|	C7 (δ)	|C7 ($\Delta\delta$)|	C5 (δ)	|C5 ($\Delta\delta$)|
0.000	6.928988	0.000000	6.530936	0.000000	6.299460	0.000000
0.054	6.926669	0.002318	6.523183	0.007753	6.293003	0.006457
0.154	6.922330	0.006658	6.508887	0.022049	6.280389	0.019071
0.246	6.918376	0.010611	6.497099	0.033836	6.269656	0.029804
0.448	6.908774	0.020213	6.470081	0.060854	6.246120	0.053340
0.616	6.900862	0.028125	6.448732	0.082204	6.227136	0.072324
0.758	6.894088	0.034900	6.430975	0.099961	6.211585	0.087875
0.986	6.882481	0.046506	6.402263	0.128673	6.186309	0.113151
1.160	6.873692	0.055296	6.381329	0.149607	6.167752	0.131708
1.297	6.866636	0.062352	6.364942	0.165994	6.153355	0.146105
1.540	6.853801	0.075186	6.335646	0.195289	6.127476	0.171984
1.699	6.845524	0.083464	6.317741	0.213195	6.111231	0.188229

All shifts in ppm

LiI to BN-Scaffold 2.48: Titration 3



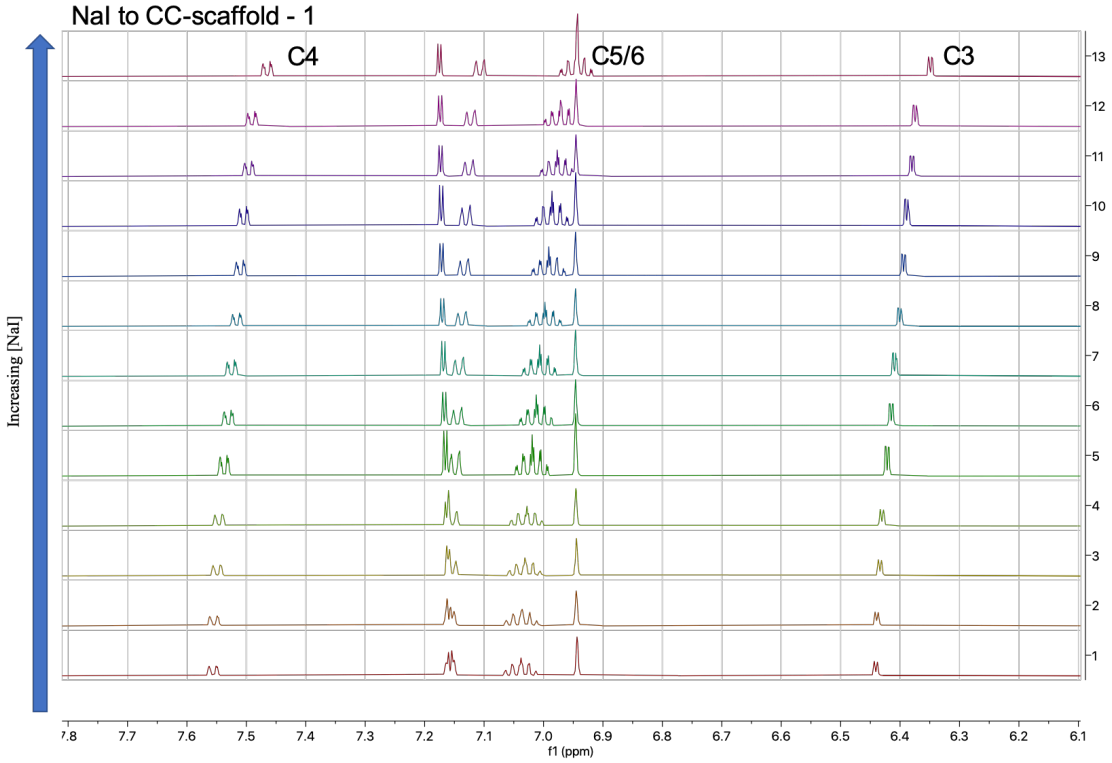
[LiI] Stock (M) **[BN-scaffold] (M)**

2.501 0.00098

[LiI] (M)	C3 (δ)	|C3 ($\Delta\delta$)|	C7 (δ)	|C7 ($\Delta\delta$)|	C5 (δ)	|C5 ($\Delta\delta$)|
0.000	6.929349	0.000000	6.531114	0.000000	6.300157	0.000000
0.054	6.926989	0.002360	6.523922	0.007193	6.293655	0.006502
0.156	6.922320	0.007029	6.509818	0.021297	6.281220	0.018937
0.250	6.918140	0.011209	6.497273	0.033842	6.270007	0.030150
0.455	6.909009	0.020340	6.471090	0.060025	6.246890	0.053267
0.625	6.901063	0.028287	6.449816	0.081298	6.228266	0.071891
0.770	6.893967	0.035382	6.431481	0.099634	6.212164	0.087993
1.001	6.882761	0.046589	6.403234	0.127880	6.187138	0.113019
1.177	6.873821	0.055529	6.381881	0.149233	6.168488	0.131669
1.316	6.866949	0.062400	6.365585	0.165530	6.153898	0.146259
1.563	6.854445	0.074904	6.337271	0.193844	6.128878	0.171279
1.725	6.845804	0.083545	6.318689	0.212425	6.111986	0.188171

All shifts in ppm

NaI to CC-Scaffold 2.49: Titration 1

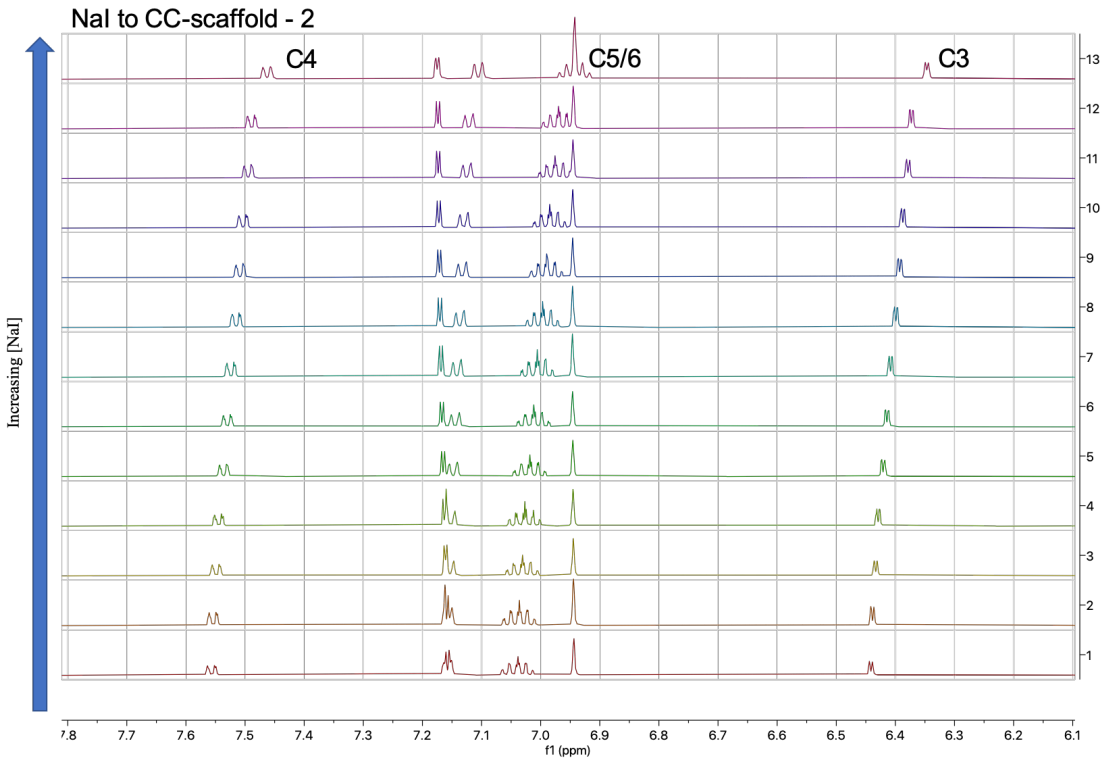


[NaI] Stock (M)	[CC-scaffold] (M)
0.817	0.00103

[NaI] (M)	C4 (δ)	|C4 ($\Delta\delta$)|	C5/C6 (δ)	|C5/C6 ($\Delta\delta$)|	C3 (δ)	|C3 ($\Delta\delta$)|
0.000	7.556692	0.000000	7.038616	0.000000	6.440964	0.000000
0.018	7.553741	0.002951	7.035894	0.002722	6.438516	0.002448
0.043	7.549074	0.007618	7.030844	0.007772	6.433243	0.007722
0.082	7.544948	0.011744	7.027023	0.011593	6.428958	0.012007
0.149	7.536619	0.020072	7.018515	0.020101	6.421224	0.019740
0.204	7.529563	0.027128	7.011399	0.027217	6.414054	0.026911
0.251	7.524023	0.032669	7.005725	0.032891	6.408427	0.032537
0.327	7.514919	0.041773	6.996480	0.042136	6.399242	0.041722
0.384	7.508311	0.048381	6.989549	0.049067	6.392515	0.048450
0.430	7.502718	0.053974	6.984248	0.054368	6.387302	0.053662
0.511	7.494282	0.062410	6.975378	0.063238	6.378562	0.062402
0.563	7.488658	0.068033	6.969324	0.069292	6.372789	0.068176
0.817	7.464801	0.091891	6.945099	0.093517	6.348925	0.092040

All shifts in ppm

NaI to CC-Scaffold 2.49: Titration 2

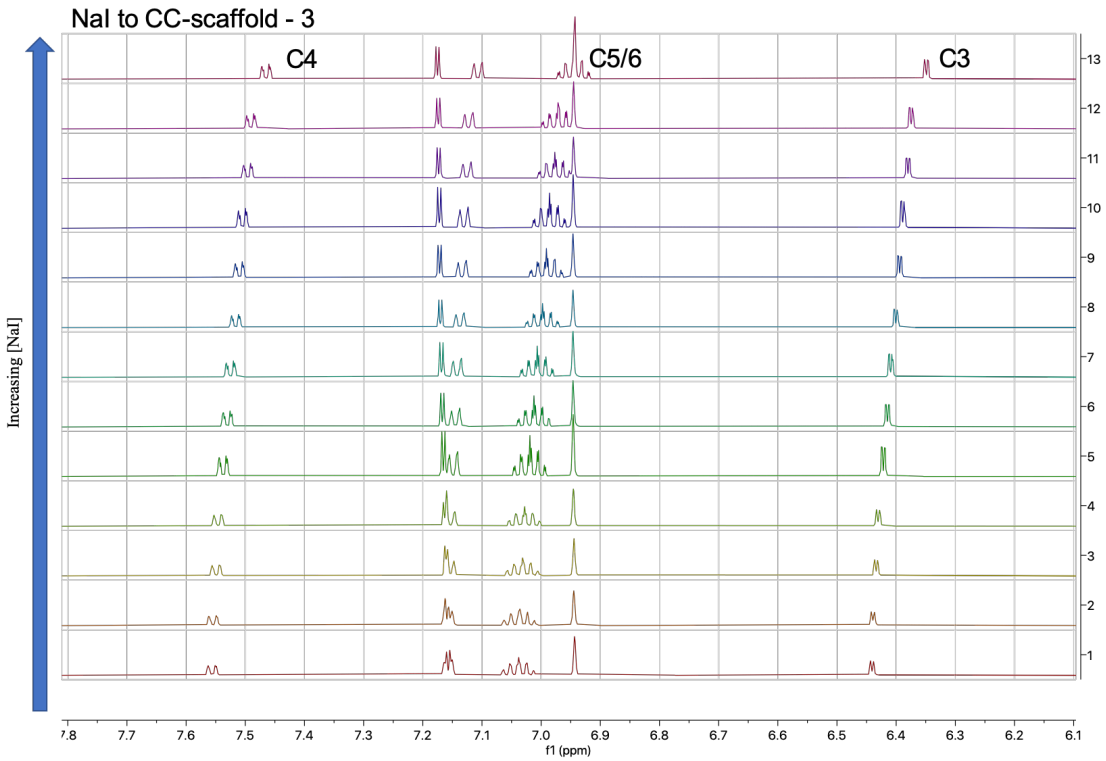


[NaI] Stock (M)	[CC-scaffold] (M)
0.820	0.00102

[NaI] (M)	C4 (δ)	|C4 ($\Delta\delta$)|	C5/C6 (δ)	|C5/C6 ($\Delta\delta$)|	C3 (δ)	|C3 ($\Delta\delta$)|
0.000	7.556859	0.000000	7.038562	0.000000	6.440914	0.000000
0.018	7.553879	0.002981	7.035912	0.002650	6.438439	0.002475
0.051	7.548905	0.007954	7.030920	0.007642	6.433603	0.007311
0.082	7.544802	0.012057	7.026828	0.011734	6.429425	0.011489
0.149	7.536514	0.020345	7.018295	0.020267	6.421003	0.019910
0.205	7.529892	0.026967	7.011651	0.026911	6.414357	0.026556
0.252	7.524127	0.032733	7.005728	0.032834	6.408410	0.032503
0.328	7.515195	0.041664	6.996671	0.041891	6.399460	0.041453
0.386	7.508499	0.048360	6.989874	0.048688	6.392797	0.048116
0.432	7.503507	0.053352	6.984711	0.053851	6.387715	0.053198
0.513	7.494745	0.062115	6.975685	0.062877	6.378984	0.061930
0.566	7.488909	0.067950	6.969712	0.068850	6.373021	0.067892
0.820	7.462144	0.094716	6.943104	0.095459	6.346838	0.094075

All shifts in ppm

NaI to CC-Scaffold 2.49: Titration 3

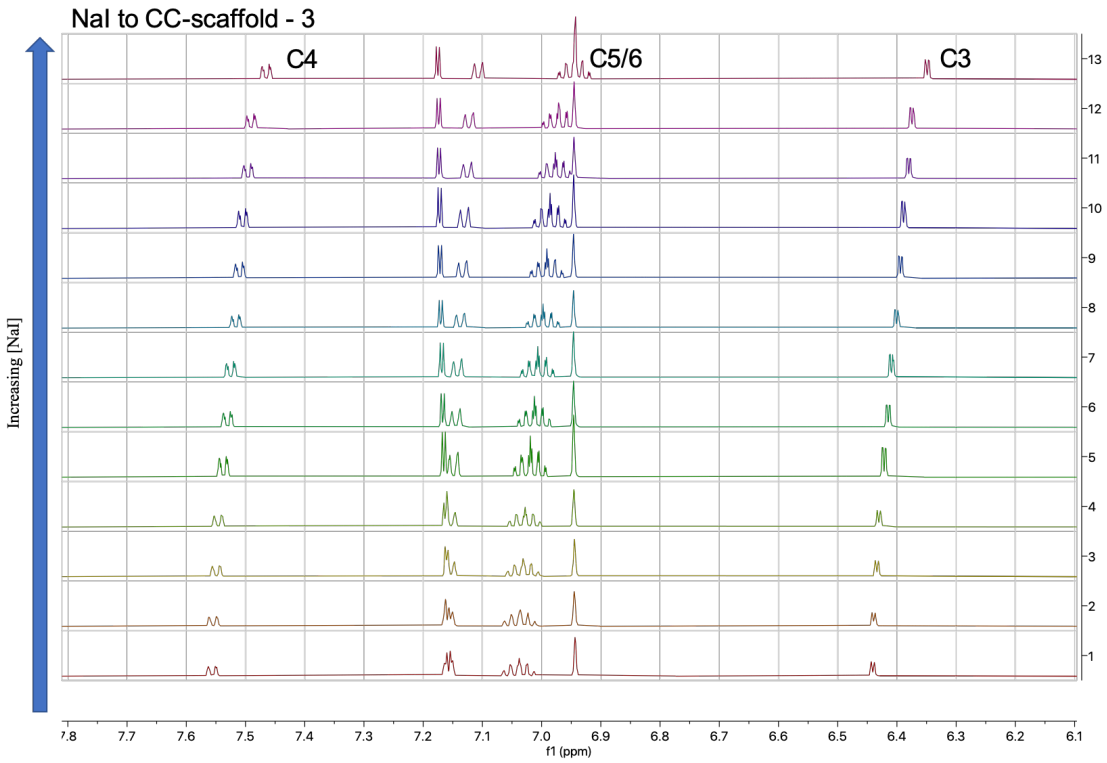


[NaI] Stock (M)	[CC-scaffold] (M)
0.792	0.00102

[NaI] (M)	C4 (δ)	|C4 ($\Delta\delta$)|	C5/C6 (δ)	|C5/C6 ($\Delta\delta$)|	C3 (δ)	|C3 ($\Delta\delta$)|
0.000	7.555884	0.000000	7.038075	0.000000	6.440790	0.000000
0.017	7.554467	0.001417	7.035628	0.002447	6.439416	0.001373
0.050	7.548906	0.006978	7.031333	0.006742	6.433832	0.006958
0.079	7.546129	0.009755	7.027799	0.010276	6.430105	0.010685
0.144	7.537226	0.018658	7.019058	0.019017	6.421726	0.019064
0.198	7.530396	0.025488	7.012196	0.025879	6.414865	0.025925
0.244	7.524996	0.030887	7.006717	0.031358	6.409299	0.031491
0.317	7.516154	0.039729	6.997659	0.040416	6.400440	0.040350
0.373	7.509789	0.046094	6.991104	0.046971	6.393986	0.046804
0.417	7.504616	0.051268	6.985786	0.052289	6.388802	0.051988
0.495	7.495922	0.059962	6.976908	0.061167	6.380030	0.060760
0.546	7.490539	0.065345	6.971452	0.066623	6.374644	0.066146
0.792	7.464801	0.091083	6.944919	0.093156	6.348929	0.091861

All shifts in ppm

NaI to BN-Scaffold 2.48: Titration 1

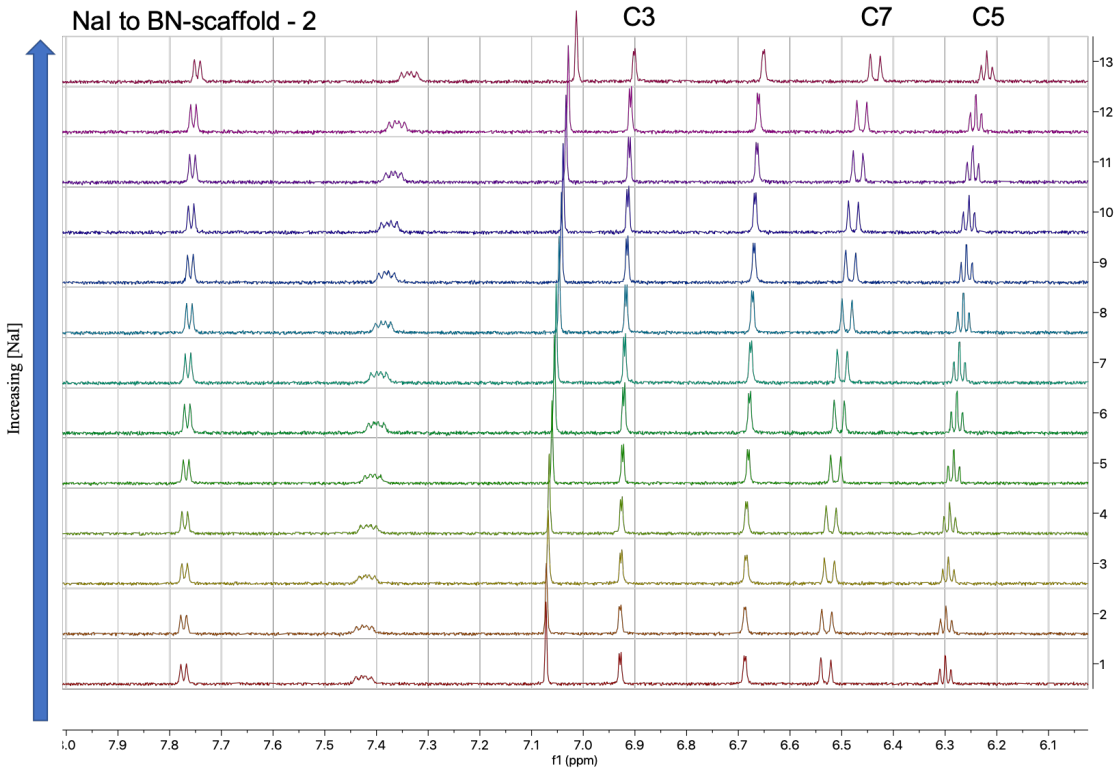


[NaI] Stock (M)	[BN-scaffold] (M)
0.771	0.00098

[NaI] (M)	C3 (δ)	|C3 ($\Delta\delta$)|	C7 (δ)	|C7 ($\Delta\delta$)|	C5 (δ)	|C5 ($\Delta\delta$)|
0.000	6.928756	0.000000	6.530705	0.000000	6.299530	0.000000
0.017	6.928639	0.000117	6.528820	0.001885	6.297635	0.001895
0.048	6.927464	0.001291	6.524584	0.006121	6.293916	0.005614
0.077	6.925962	0.002793	6.520168	0.010537	6.290828	0.008702
0.140	6.923800	0.004956	6.511990	0.018715	6.283493	0.016037
0.193	6.921708	0.007048	6.505030	0.025676	6.277690	0.021840
0.237	6.920490	0.008266	6.499470	0.031235	6.272898	0.026632
0.308	6.917447	0.011309	6.490352	0.040354	6.265115	0.034415
0.363	6.915768	0.012988	6.483770	0.046935	6.259614	0.039916
0.406	6.914096	0.014660	6.478355	0.052351	6.254976	0.044554
0.482	6.911069	0.017687	6.468905	0.061800	6.247288	0.052242
0.531	6.909434	0.019321	6.463284	0.067421	6.242625	0.056905
0.771	6.901348	0.027408	6.436263	0.094443	6.220235	0.079295

All shifts in ppm

NaI to BN-Scaffold 2.48: Titration 2

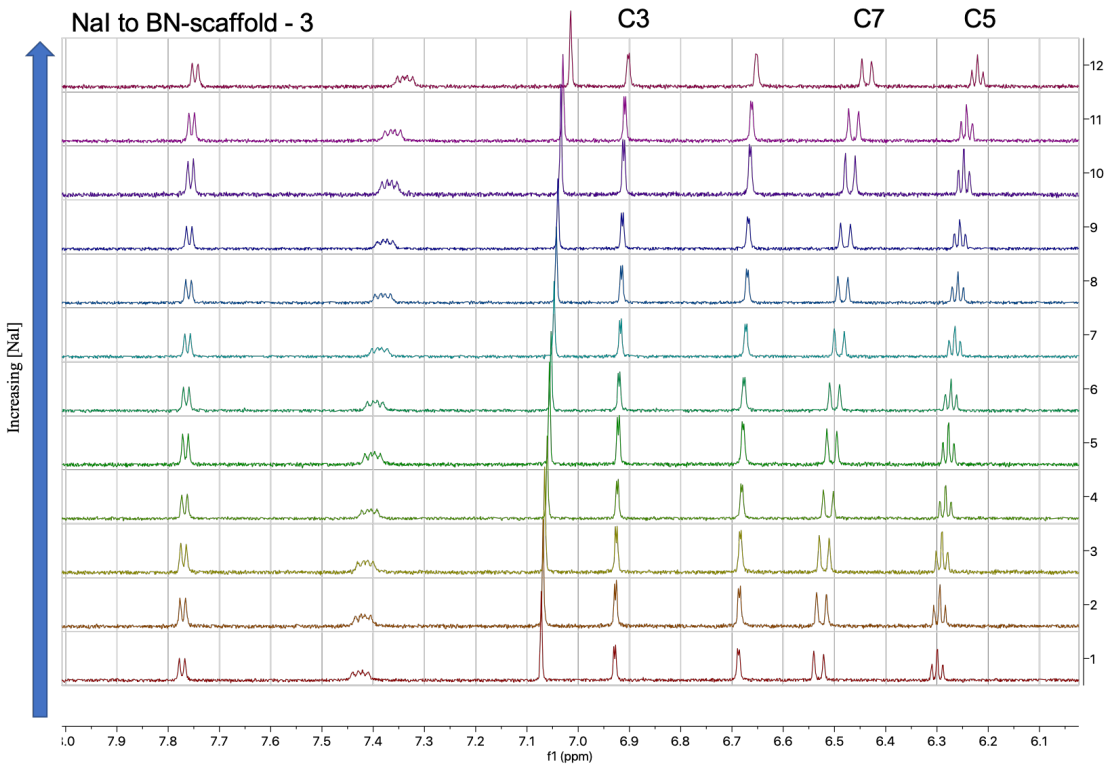


[NaI] Stock (M)	[BN-scaffold] (M)
0.771	0.00098

[NaI] (M)	C3 (δ)	|C3 ($\Delta\delta$)|	C7 (δ)	|C7 ($\Delta\delta$)|	C5 (δ)	|C5 ($\Delta\delta$)|
0.000	6.928788	0.000000	6.530669	0.000000	6.299615	0.000000
0.017	6.928938	0.000149	6.528929	0.001740	6.298209	0.001406
0.048	6.927013	0.001776	6.523737	0.006932	6.293749	0.005866
0.077	6.926443	0.002346	6.520444	0.010225	6.290672	0.008943
0.140	6.923809	0.004979	6.511702	0.018967	6.283169	0.016446
0.193	6.921539	0.007250	6.504625	0.026043	6.277288	0.022327
0.237	6.919967	0.008821	6.498878	0.031791	6.272471	0.027144
0.309	6.917291	0.011498	6.489681	0.040988	6.264531	0.035084
0.363	6.915142	0.013647	6.482632	0.048037	6.258696	0.040919
0.406	6.913821	0.014968	6.477346	0.053322	6.254433	0.045182
0.482	6.910806	0.017983	6.468244	0.062425	6.246842	0.052773
0.532	6.909033	0.019756	6.461047	0.069621	6.240732	0.058883
0.771	6.901437	0.027352	6.435029	0.095639	6.219413	0.080202

All shifts in ppm

NaI to BN-Scaffold 2.48: Titration 3



[NaI] Stock (M) 0.777

[BN-scaffold] (M) 0.00098

[NaI] (M)	C3 (δ)	|C3 ($\Delta\delta$)|	C7 (δ)	|C7 ($\Delta\delta$)|	C5 (δ)	|C5 ($\Delta\delta$)|
0.000	6.928667	0.000000	6.530696	0.000000	6.299555	0.000000
0.041	6.927420	0.001247	6.525136	0.005560	6.294802	0.004753
0.078	6.925657	0.003010	6.519908	0.010789	6.290386	0.009169
0.141	6.923779	0.004888	6.511894	0.018803	6.283563	0.015992
0.194	6.921792	0.006875	6.505005	0.025692	6.277680	0.021875
0.239	6.920274	0.008393	6.499728	0.030969	6.273091	0.026464
0.311	6.917724	0.010943	6.490563	0.040133	6.265375	0.034180
0.366	6.915478	0.013189	6.483516	0.047180	6.259588	0.039967
0.409	6.913946	0.014721	6.478370	0.052326	6.255248	0.044307
0.486	6.911329	0.017338	6.469174	0.061522	6.247577	0.051978
0.536	6.909035	0.019632	6.462515	0.068181	6.241998	0.057557
0.777	6.902349	0.026318	6.436925	0.093771	6.220953	0.078602

All shifts in ppm

2.6.6. Titration Fit Data

*CC-Scaffold **2.49** Data: Lithium Iodide*

π6-system		
Titration	**Ka (M^{-1})**	**Error (%)**
1	0.05502084	5.44686547
2	0.06391472	0.87493265
3	0.06004288	0.84534222
Average	**0.05965948**	
STDEV	**0.00445932**	

*BN-Scaffold **2.48** Data: Lithium Iodide*

π6-system		
Titration	**Ka (M^{-1})**	**Error (%)**
1	0.05143333	0.14510105
2	0.05667633	0.21559881
3	0.05673581	0.2334196
Average	**0.05494849**	
STDEV	**0.00304436**	

*CC-Scaffold **2.49** Data: Sodium Iodide*

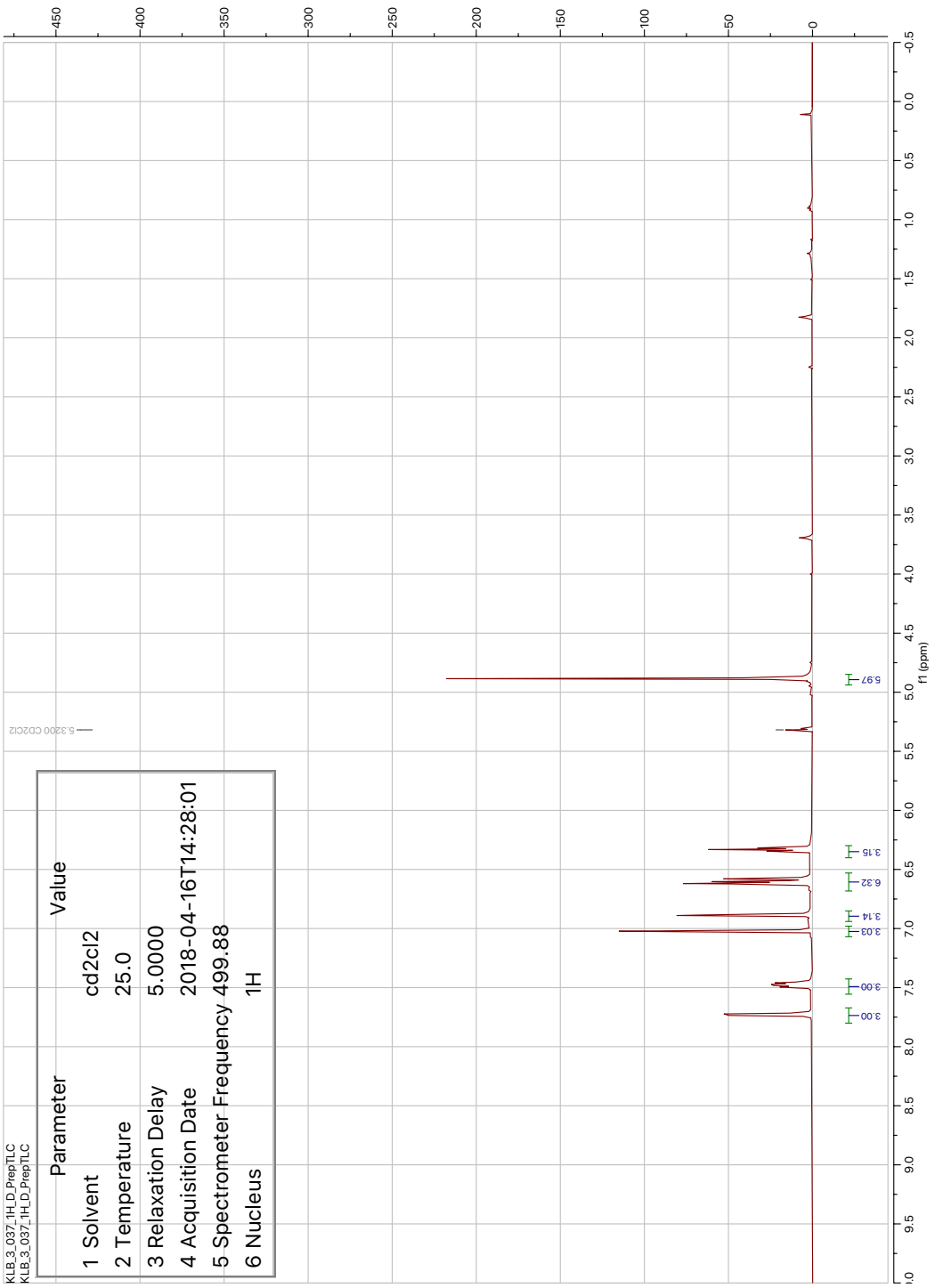
π6-system		
Titration	**Ka (M^{-1})**	**Error (%)**
1	0.26928284	1.15561696
2	0.22408539	0.56017266
3	0.18666979	0.22168182
Average	**0.22667934**	
STDEV	**0.04136757**	

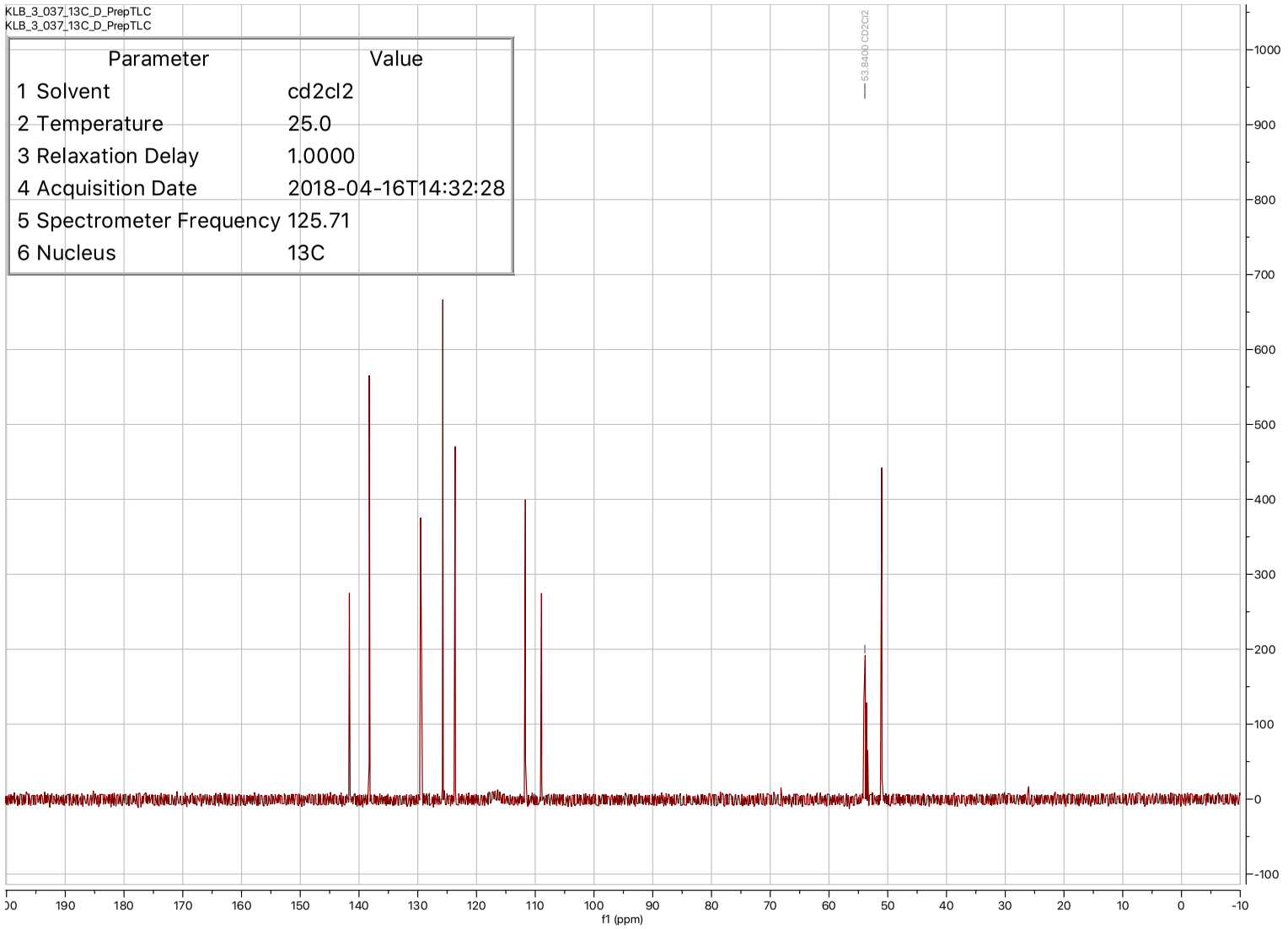
*BN-Scaffold **2.48** Data: Sodium Iodide*

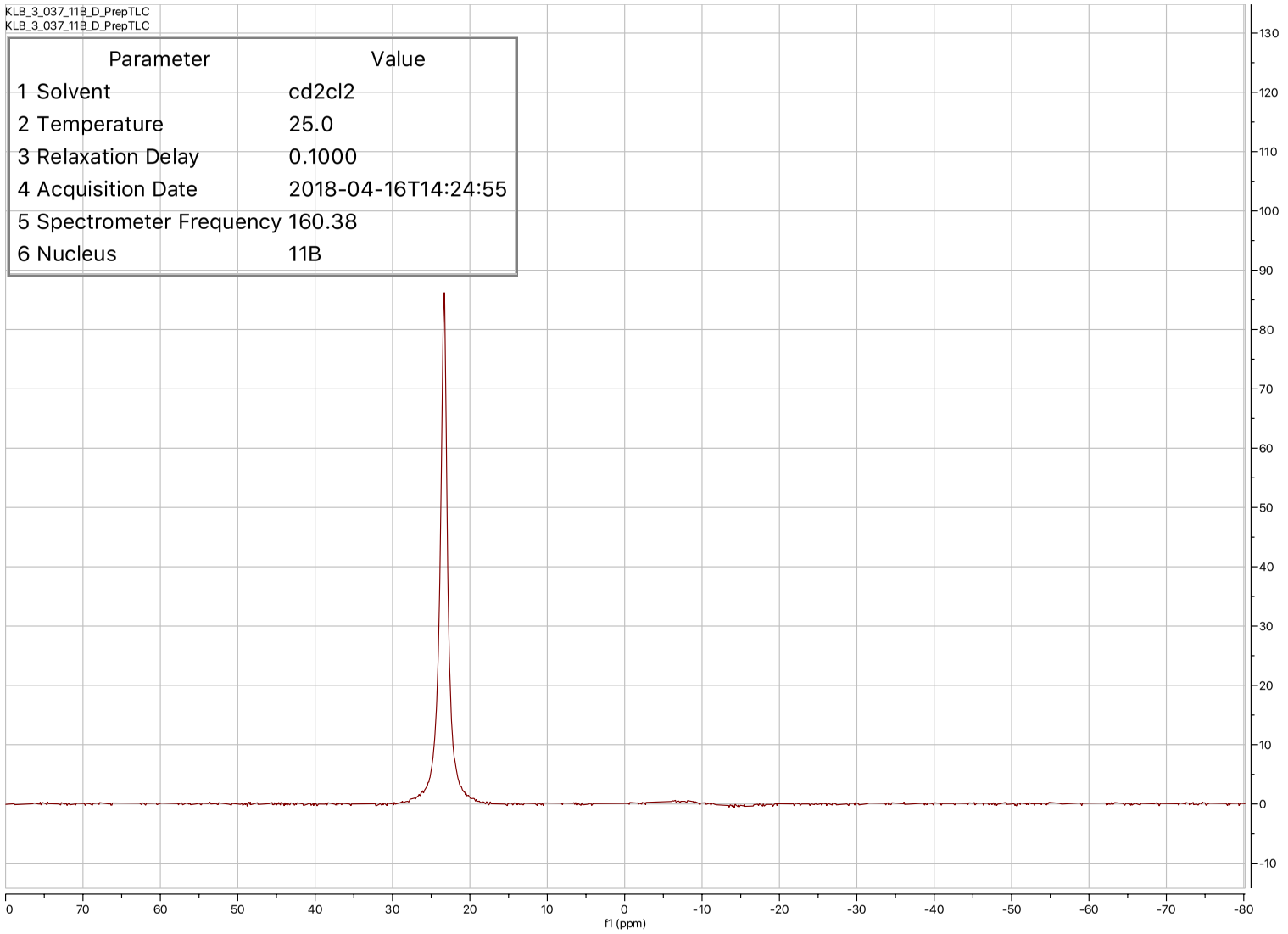
π6-system		
Titration	**Ka (M^{-1})**	**Error (%)**
1	0.18584887	0.22566179
2	0.20604828	0.67987845
3	0.20170994	0.56146666
Average	**0.19786903**	
STDEV	**0.01063337**	

2.6.7. NMR Spectra

BN-Indole Scaffold 2.48




KLB_3_037_13C_D_PrepTLC
KLB_3_037_13C_D_PrepTLC
Parameter Value
1 Solvent cd2cl2
2 Temperature 25.0
3 Relaxation Delay 1.0000
4 Acquisition Date 2018-04-16T14:32:28
5 Spectrometer Frequency 125.71
6 Nucleus 13C
53.840 CD2Cl2
f1 (ppm)


KLB_3_037_11B_D_PrepTLC
KLB_3_037_11B_D_PrepTLC
Parameter Value
1 Solvent cd2cl2
2 Temperature 25.0
3 Relaxation Delay 0.1000
4 Acquisition Date 2018-04-16T14:24:55
5 Spectrometer Frequency 160.38
6 Nucleus 11B
f1 (ppm)

2.6.8. Computational Results

1. Indole

Neutral, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.000000	0.711232	0.000000
2	6	0.490090	-0.612318	0.000000
3	6	-0.425045	-1.671507	0.000000
4	6	-1.775002	-1.391529	0.000000
5	6	-2.240249	-0.066457	0.000000
6	6	-1.363946	0.998836	0.000000
7	6	2.240895	0.795595	0.000000
8	6	1.922227	-0.526167	0.000000
9	1	-0.075004	-2.697298	0.000000
10	1	-2.492461	-2.203204	0.000000
11	1	-3.307158	0.121632	0.000000
12	1	-1.725094	2.020807	0.000000
13	1	1.052471	2.553920	0.000000
14	1	3.207667	1.273508	0.000000
15	1	2.622146	-1.346159	0.000000
16	7	1.089088	1.550098	0.000000

SCF Energy of neutral: -363.7026772610000 au
Dipole moment (Debye):
x= -0.876 y= 1.993 z= 0.000 Norm= 2.177

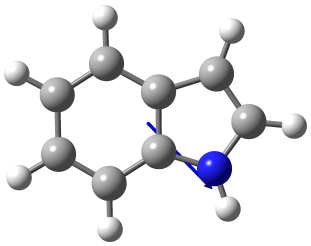
HOMO = -7.052 eV LUMO = 0.885 eV gap(HOMO-LUMO) = 7.937 eV

MO	(#)	au	eV
LUMO+4	(36)	0.11318	3.0798
LUMO+3	(35)	0.10061	2.7377
LUMO+2	(34)	0.07899	2.1494
LUMO+1	(33)	0.07229	1.9671
LUMO	(32)	0.03252	0.8849
HOMO	(31)	-0.25916	-7.0521
HOMO-1	(30)	-0.27574	-7.5033
HOMO-2	(29)	-0.34030	-9.2600
HOMO-3	(28)	-0.39785	-10.8260
HOMO-4	(27)	-0.40347	-10.9790
HOMO-5	(26)	-0.41931	-11.4100
HOMO-6	(25)	-0.45685	-12.4315
HOMO-7	(24)	-0.47925	-13.0411
HOMO-8	(23)	-0.49597	-13.4960
HOMO-9	(22)	-0.49679	-13.5183
HOMO-10	(21)	-0.50561	-13.7583
HOMO-11	(20)	-0.54120	-14.7268

Gap (au ; eV ; nm) 0.29168 ; 7.9370 ; 156.21

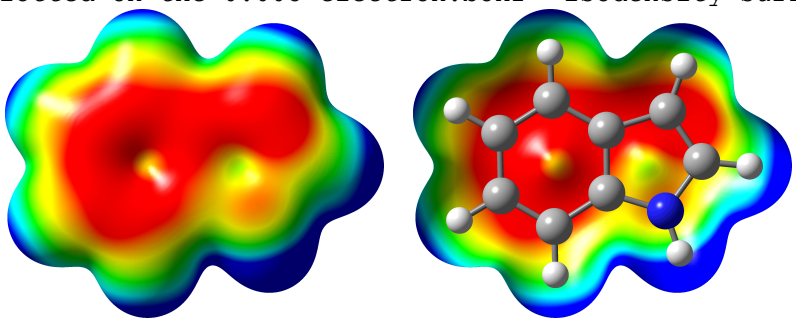
Zero-point correction= 0.130909
(Hartree/Particle)
Sum of electronic and zero-point Energies= -363.571768 au

Dipole moment (debye):

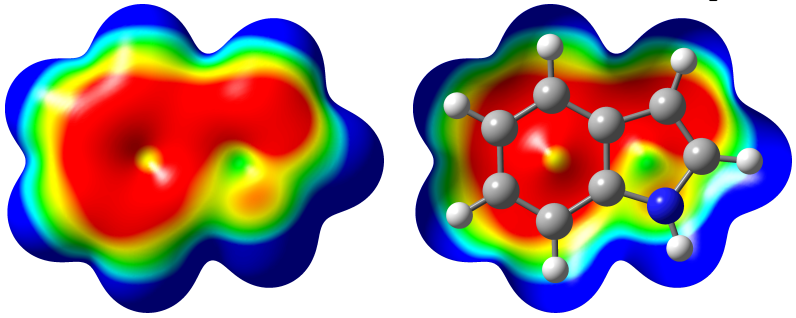


2.177 D

Electrostatic potential map (from -13.81 (red) to +27.61 (bleu) kcal/mol) plotted on the 0.005 electron.bohr^{-3} isodensity surface.



Electrostatic potential map (from -12.55 (red) to +12.55 (bleu) kcal/mol) plotted on the 0.005 electron.bohr^{-3} isodensity surface.



The number of surface minima: 8

#	a.u.	eV	kcal/mol	X/Y/Z coordinate(Angstrom)		
1	-0.03518150	-0.957337	-22.076743	-2.066262	-0.128122	-1.611475
2	-0.03518728	-0.957495	-22.080373	-2.087543	-0.139297	1.610904
3	-0.01561487	-0.424902	-9.798488	-1.115165	1.378661	-1.534010
4	-0.01562173	-0.425089	-9.802791	-1.133959	1.368612	1.532804
5	-0.03657775	-0.995331	-22.952907	1.107148	-1.320415	1.561307
6	-0.03659599	-0.995828	-22.964349	1.166566	-1.331020	-1.565691
7	-0.03356317	-0.913300	-21.061224	1.723934	0.258029	-1.563103
8	-0.03357063	-0.913503	-21.065907	1.658292	0.263506	1.555075

Location of minima (ESP in kcal/mol) on 0.005 electron.bohr^{-3} isodensity surface:

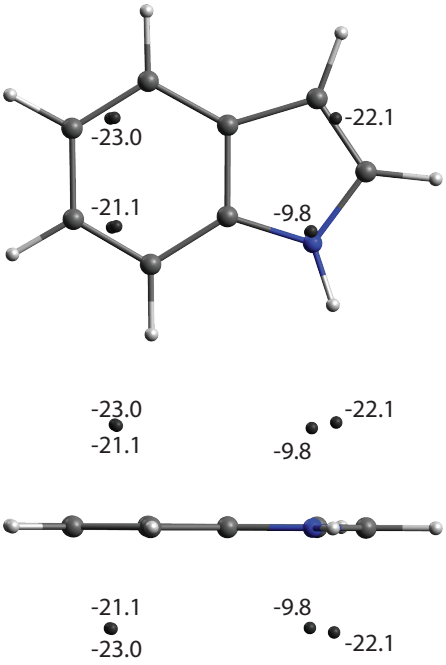
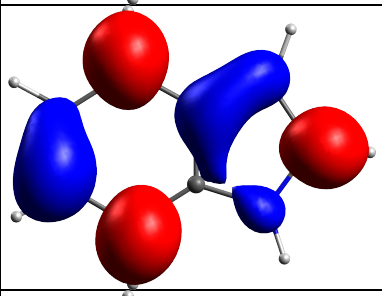
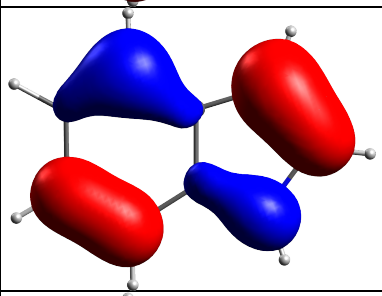
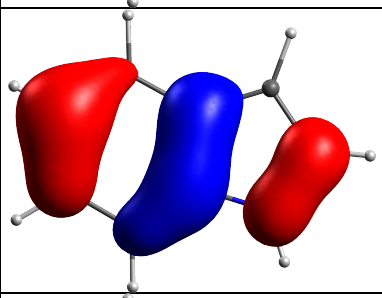
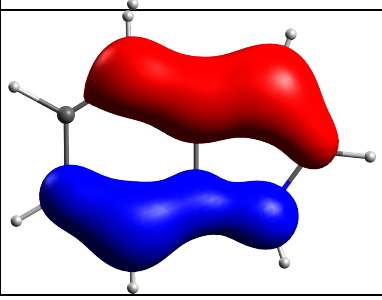
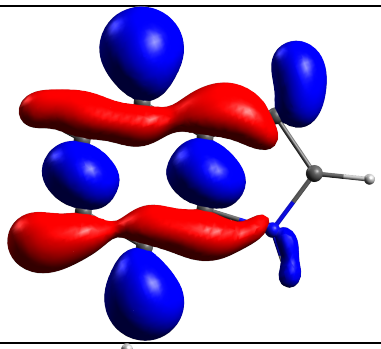
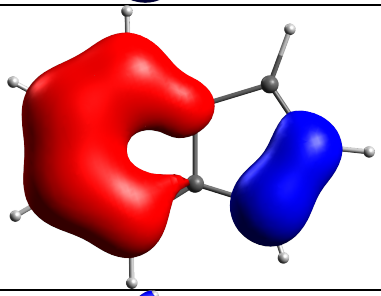
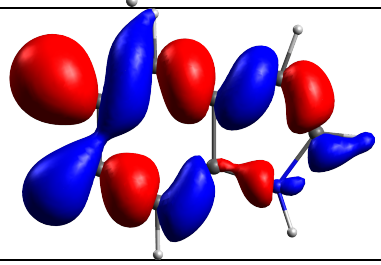

-23.0
-22.1
-21.1
-9.8
-23.0
-22.1
-21.1
-9.8
-21.1
-9.8
-23.0
-22.1

Table: molecular orbitals, energies

# MO		CAM-B3LYP ε^{KS} (eV)
32 (LUMO)		0.8849
31 (HOMO)		-7.0521
30		-7.5033
29		-9.2600

28		-10.8260
27		-10.9790
26		-11.4100

2. BN-indole

Neutral, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	-1.334244	0.978266	0.000000
2	6	-2.227904	-0.044215	0.000000
3	6	-1.798200	-1.406225	0.000000
4	6	-0.481910	-1.769279	0.000000
5	6	2.217629	0.892290	0.000000
6	6	1.072028	1.604113	0.000000
7	1	-1.633845	2.019812	0.000000
8	1	-3.284938	0.185787	0.000000
9	1	-2.573381	-2.167456	0.000000
10	1	-0.236146	-2.825698	0.000000
11	1	2.679696	-1.148344	0.000000
12	1	3.226789	1.272740	0.000000
13	1	0.933045	2.672752	0.000000
14	7	1.943389	-0.466079	0.000000
15	5	0.520134	-0.648133	0.000000
16	7	0.000000	0.709132	0.000000

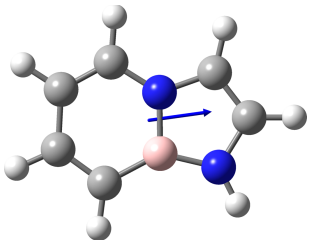
SCF Energy of neutral: -367.1610610100000 au
Dipole moment (Debye):
x= 1.332 y= 0.716 z= 0.000 Norm= 1.512

HOMO = -7.061 eV LUMO = 0.742 eV gap(HOMO-LUMO) = 7.803 eV

MO	(#)	au	eV
LUMO+4	(36)	0.11263	3.0648
LUMO+3	(35)	0.09827	2.6741
LUMO+2	(34)	0.08759	2.3834
LUMO+1	(33)	0.08359	2.2746
LUMO	(32)	0.02725	0.7415
HOMO	(31)	-0.25950	-7.0614
HOMO-1	(30)	-0.26723	-7.2717
HOMO-2	(29)	-0.35577	-9.6810
HOMO-3	(28)	-0.38222	-10.4007
HOMO-4	(27)	-0.40873	-11.1221
HOMO-5	(26)	-0.42471	-11.5569
HOMO-6	(25)	-0.45668	-12.4269
HOMO-7	(24)	-0.47925	-13.0411
HOMO-8	(23)	-0.49473	-13.4623
HOMO-9	(22)	-0.49523	-13.4759
HOMO-10	(21)	-0.51047	-13.8906
HOMO-11	(20)	-0.53036	-14.4318
Gap (au ; eV ; nm)		0.28675 ;	7.8029 ;	158.90

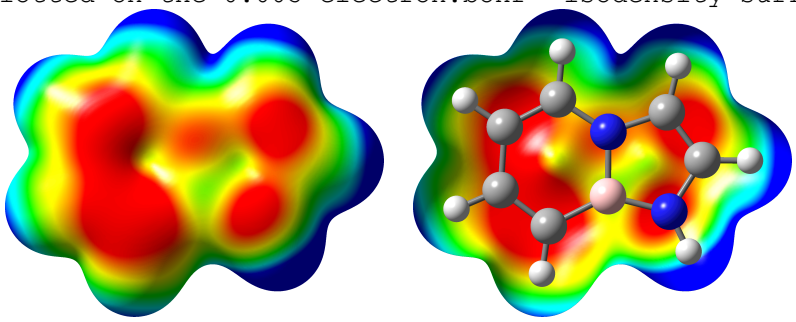
Zero-point correction= 0.129295 (Hartree/Particle)
Sum of electronic and zero-point Energies= -367.031766 au

Dipole moment (debye):

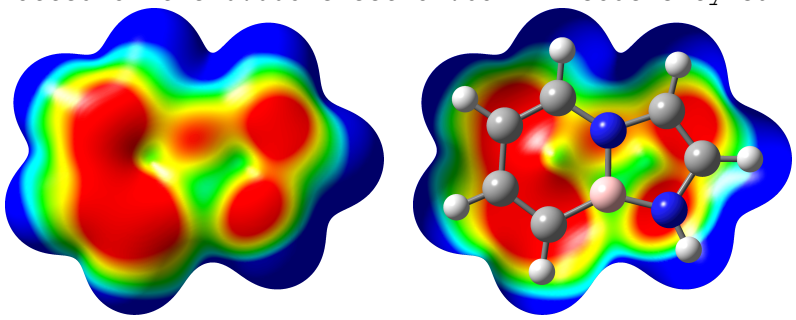


1.512 D

Electrostatic potential map (from -13.81 (red) to +27.61 (bleu) kcal/mol) plotted on the 0.005 electron.bohr^{-3} isodensity surface.



Electrostatic potential map (from −12.55 (red) to +12.55 (bleu) kcal/mol) plotted on the 0.005 electron.bohr^{-3} isodensity surface.



The number of surface minima: 10

#	a.u.	eV	kcal/mol	X/Y/Z coordinate(Angstrom)		
1	-0.03244464	-0.882864	-20.359338	-1.901013	0.143695	-1.594575
2	-0.03245238	-0.883074	-20.364192	-1.891513	0.152396	1.593999
3	-0.04016062	-1.092826	-25.201189	-0.951047	-1.524881	-1.597660
4	-0.04014069	-1.092284	-25.188683	-0.936221	-1.513266	1.597074
5	-0.01938094	-0.527382	-12.161737	-0.082306	0.455646	-1.524086
6	-0.01939535	-0.527774	-12.170773	-0.086083	0.472151	1.524952
7	-0.03081800	-0.838601	-19.338606	1.641866	1.446796	1.608209
8	-0.03080133	-0.838147	-19.328141	1.665684	1.407303	-1.608720
9	-0.02870841	-0.781196	-18.014814	1.799702	-0.499986	-1.567872
10	-0.02869120	-0.780727	-18.004013	1.768537	-0.550404	1.566465

Location of minima (ESP in kcal/mol) on 0.005 electron.bohr^{-3} isodensity surface:

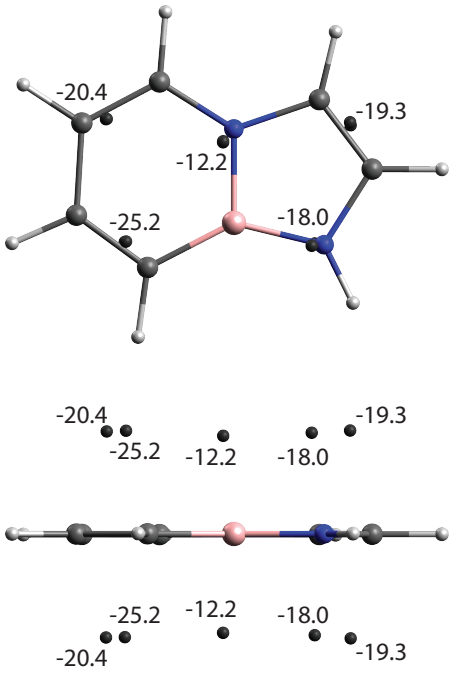
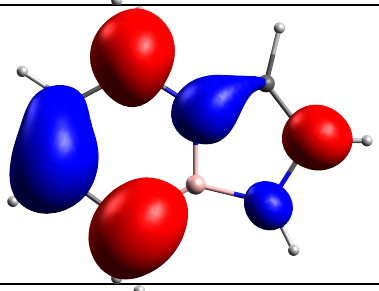
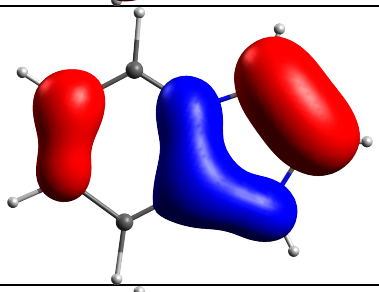
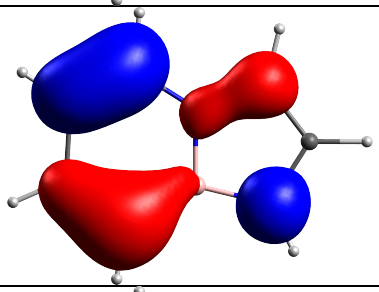
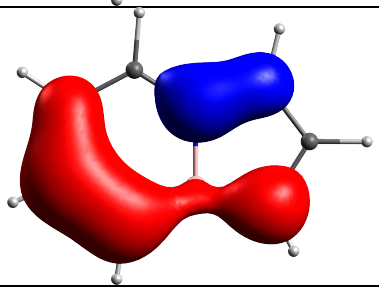
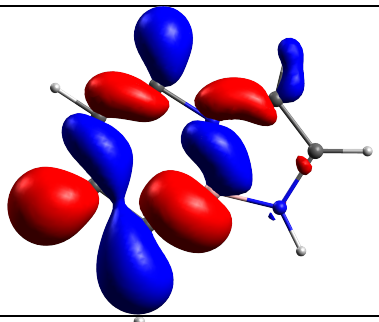
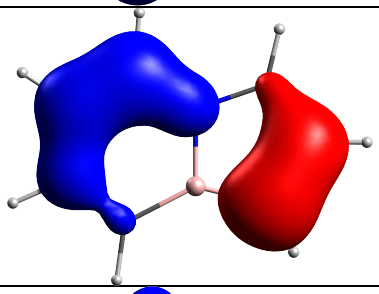
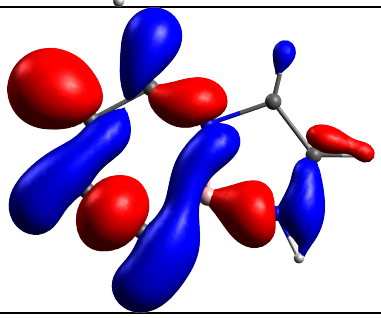

-20.4
-19.3
-12.2
-25.2
-18.0
-20.4
-19.3
-25.2
-12.2
-18.0
-25.2
-12.2
-18.0
-20.4
-19.3

Table: molecular orbitals, energies

# MO		CAM-B3LYP ε^{KS} (eV)
32 (LUMO)		0.7415
31 (HOMO)		-7.0614
30		-7.2717
29		-9.6810

28		-10.4007
27		-11.1221
26		-11.5569

3. Cation Li^+

Cation, ground state, CAM-B3LYP/6-311G(d)

```
SCF Energy of neutral:    -7.264261449600000       au

HOMO =   -65.857 eV   LUMO =    -6.324 eV   gap(HOMO-LUMO) =    59.533
eV

  MO    (#)          au          eV
LUMO+4  (  6)   -0.06765    -1.8409
LUMO+3  (  5)   -0.15880    -4.3212
LUMO+2  (  4)   -0.15880    -4.3212
LUMO+1  (  3)   -0.15880    -4.3212
LUMO    (  2)   -0.23242    -6.3245
HOMO    (  1)   -2.42020   -65.8570
```

4. Cation Na^+

Cation, ground state, CAM-B3LYP/6-311G(d)

```
SCF Energy of neutral:    -162.0643713740000       au

HOMO =   -41.191 eV   LUMO =    -6.287 eV   gap(HOMO-LUMO) =    34.904
eV

  MO    (#)          au          eV
LUMO+4  ( 10)   -0.06495    -1.7674
LUMO+3  (  9)   -0.13293    -3.6172
LUMO+2  (  8)   -0.13293    -3.6172
LUMO+1  (  7)   -0.13293    -3.6172
LUMO    (  6)   -0.23106    -6.2875
HOMO    (  5)   -1.51374   -41.1910
HOMO-1  (  4)   -1.51374   -41.1910
HOMO-2  (  3)   -1.51374   -41.1910
HOMO-3  (  2)   -2.57941   -70.1893
HOMO-4  (  1)  -38.82095 -1056.3718
```

5. Cation K^+

Cation, ground state, CAM-B3LYP/6-311G(d)

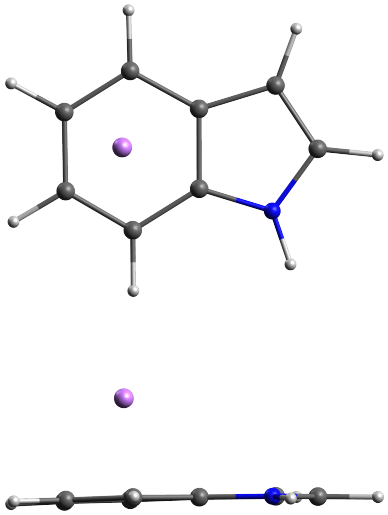
```
SCF Energy of neutral:    -599.7694236890000       au

HOMO =   -28.418 eV   LUMO =    -5.134 eV   gap(HOMO-LUMO) =    23.284
eV

  MO    (#)          au          eV
LUMO+4  ( 14)   -0.05987    -1.6291
LUMO+3  ( 13)   -0.11632    -3.1652
LUMO+2  ( 12)   -0.11632    -3.1652
LUMO+1  ( 11)   -0.11632    -3.1652
LUMO    ( 10)   -0.18867    -5.1340
HOMO    (  9)   -1.04433   -28.4177
HOMO-1  (  8)   -1.04433   -28.4177
HOMO-2  (  7)   -1.04433   -28.4177
HOMO-3  (  6)   -1.68873   -45.9527
```

```
HOMO-4  (  5)  -10.85365  -295.3429
HOMO-5  (  4)  -10.85365  -295.3429
HOMO-6  (  3)  -10.85365  -295.3429
HOMO-7  (  2)  -13.50678  -367.5382
HOMO-8  (  1) -130.08719 -3539.8526
```

6. Indole-π_6...Li^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.307921	-0.680635	-0.062389
2	6	0.304920	0.746816	-0.058240
3	6	-0.926799	1.423339	-0.118133
4	6	-2.103741	0.688059	-0.168964
5	6	-2.080104	-0.722875	-0.149710
6	6	-0.880108	-1.421052	-0.088830
7	6	2.426148	0.032177	-0.017362
8	6	1.672590	1.164162	-0.019563
9	1	-0.955863	2.506025	-0.134369
10	1	-3.055383	1.200282	-0.228964
11	1	-3.012514	-1.271050	-0.189815
12	1	-0.870532	-2.504342	-0.084801
13	1	1.939746	-2.031280	-0.070010
14	1	3.499091	-0.078188	-0.009933
15	1	2.043544	2.175720	-0.020285
16	7	1.613903	-1.078317	-0.028818
17	3	-1.070125	0.057036	1.679682



SCF Energy of neutral: -371.0455579090000 au
Dipole moment (Debye):
x= -1.279 y= -1.384 z= 4.479 Norm= 4.859

HOMO = -11.835 eV LUMO = -4.188 eV gap(HOMO-LUMO) = 7.647 eV

MO	(#)	au	eV
LUMO+4	(37)	-0.08743	-2.3791
LUMO+3	(36)	-0.09173	-2.4961
LUMO+2	(35)	-0.11866	-3.2289
LUMO+1	(34)	-0.14788	-4.0240
LUMO	(33)	-0.15390	-4.1878

HOMO (32) -0.43491 -11.8345
HOMO-1 (31) -0.46706 -12.7093
HOMO-2 (30) -0.52901 -14.3951
Gap (au ; eV ; nm) 0.28101 ; 7.6467 ; 162.14

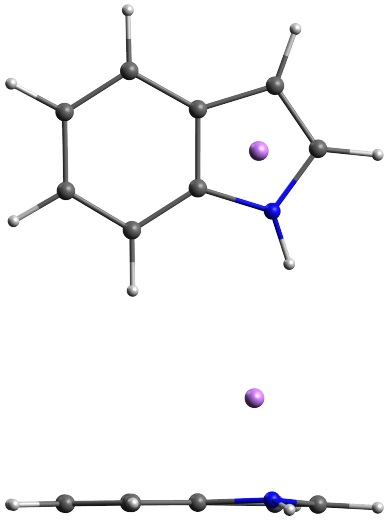
Zero-point correction= 0.134049
(Hartree/Particle)
Sum of electronic and zero-point Energies= -370.909818 au

Counterpoise fragments' energies:
Indole -363.703468307 au
Li^+ -7.26441810962 au
Isolated fragments' energies at complex geometry:
Indole -363.701934411 au
Li^+ -7.26426144960 au

7. Indole-π_5...Li^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.177993	-0.669158	-0.084717
2	6	0.180179	0.742733	-0.080471
3	6	-1.048582	1.428923	-0.030936
4	6	-2.207140	0.692550	-0.002892
5	6	-2.183485	-0.717010	-0.013880
6	6	-1.001999	-1.420371	-0.047657
7	6	2.329134	0.036172	-0.203107
8	6	1.559376	1.163833	-0.152531
9	1	-1.074941	2.511234	-0.036479
10	1	-3.162953	1.199727	0.020717
11	1	-3.120663	-1.258601	0.000025
12	1	-0.988867	-2.502872	-0.066521
13	1	1.814569	-2.023724	-0.288158
14	1	3.397941	-0.073405	-0.304481
15	1	1.926161	2.177323	-0.203503
16	7	1.509795	-1.080657	-0.092047
17	3	1.269108	-0.003707	1.739957



SCF Energy of neutral: -371.0372673710000 au
Dipole moment (Debye):
x= 4.458 y= -1.368 z= 4.578 Norm= 6.534

HOMO = -11.942 eV LUMO = -4.276 eV gap(HOMO-LUMO) = 7.666 eV

MO	(#)	au	eV
LUMO+4	(37)	-0.09216	-2.5078
LUMO+3	(36)	-0.09978	-2.7152
LUMO+2	(35)	-0.10507	-2.8591
LUMO+1	(34)	-0.14184	-3.8597

```
  LUMO    ( 33)  -0.15713   -4.2757
  HOMO    ( 32)  -0.43885  -11.9417
  HOMO-1  ( 31)  -0.45798  -12.4623
  HOMO-2  ( 30)  -0.52813  -14.3711
Gap (au ; eV ; nm)        0.28172 ;    7.6660 ;    161.73

Zero-point correction=                          0.133732
(Hartree/Particle)
Sum of electronic and zero-point Energies=        -370.901519 au
```

Counterpoise fragments' energies:

Indole -363.703079724 au

Li^+ -7.26441400449 au

Isolated fragments' energies at complex geometry:

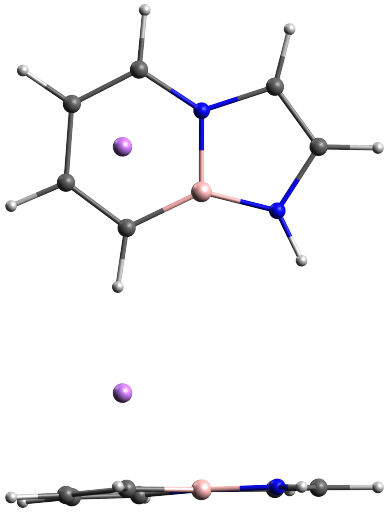
Indole -363.701215986 au

Li^+ -7.26426144960 au

8. BN-indole-π_6...Li^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.840229	-1.397821	-0.113840
2	6	2.051955	-0.756203	-0.167598
3	6	2.139364	0.671872	-0.170467
4	6	1.026588	1.478399	-0.083523
5	6	-2.440850	-0.038990	-0.019109
6	6	-1.633256	-1.119981	-0.009866
7	1	0.750872	-2.477666	-0.128804
8	1	2.952023	-1.351542	-0.236550
9	1	3.130216	1.109654	-0.227659
10	1	1.172726	2.552271	-0.075182
11	1	-2.158442	2.030582	-0.065137
12	1	-3.519508	-0.033258	-0.013545
13	1	-1.885926	-2.167071	-0.005600
14	7	-1.706480	1.131192	-0.039471
15	5	-0.323394	0.805641	-0.081230
16	7	-0.306788	-0.665544	-0.035180
17	3	1.121236	0.008546	1.689203



SCF Energy of neutral: -374.4998996900000 au
Dipole moment (Debye):
x= 1.487 y= -0.161 z= 4.490 Norm= 4.732

HOMO = -11.657 eV LUMO = -4.521 eV gap(HOMO-LUMO) = 7.136 eV

MO	(#)	au	eV
LUMO+4	(37)	-0.08731	-2.3758
LUMO+3	(36)	-0.09046	-2.4615
LUMO+2	(35)	-0.10247	-2.7884
LUMO+1	(34)	-0.14908	-4.0567
LUMO	(33)	-0.16615	-4.5212

HOMO (32) -0.42840 -11.6574
HOMO-1 (31) -0.46040 -12.5281
HOMO-2 (30) -0.54588 -14.8542
Gap (au ; eV ; nm) 0.26225 ; 7.1362 ; 173.74

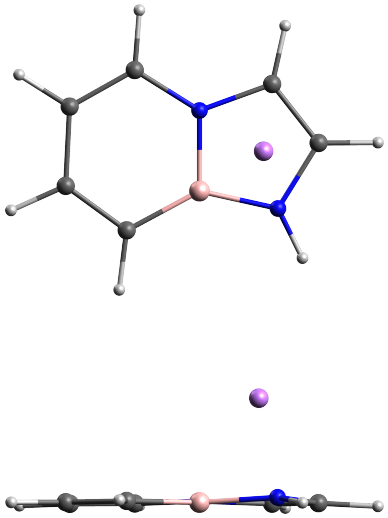
Zero-point correction= 0.132230
(Hartree/Particle)
Sum of electronic and zero-point Energies= -374.366043 au

Counterpoise fragments' energies:
BN-indole -367.161061882 au
Li^+ -7.26443319399 au
Isolated fragments' energies at complex geometry:
BN-indole -367.159606133 au
Li^+ -7.26426144960 au

9. BN-indole-π_5...Li^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.972560	-1.406347	-0.025133
2	6	2.157978	-0.759109	0.002985
3	6	2.245468	0.671083	-0.005855
4	6	1.152158	1.486152	-0.056226
5	6	-2.338602	-0.050831	-0.206959
6	6	-1.511313	-1.124681	-0.147662
7	1	0.876269	-2.484231	-0.031727
8	1	3.066400	-1.345310	0.027378
9	1	3.240879	1.101336	0.022957
10	1	1.299045	2.559158	-0.068694
11	1	-2.022724	2.024134	-0.247109
12	1	-3.412570	-0.044716	-0.312327
13	1	-1.758255	-2.173447	-0.204266
14	7	-1.589859	1.123983	-0.102888
15	5	-0.173321	0.793466	-0.123510
16	7	-0.190863	-0.661523	-0.057462
17	3	-1.342296	0.086973	1.728963



SCF Energy of neutral: -374.4936479500000 au
Dipole moment (Debye):
x= -4.828 y= -0.026 z= 4.522 Norm= 6.615

HOMO = -11.847 eV LUMO = -4.322 eV gap(HOMO-LUMO) = 7.525 eV

MO	(#)	au	eV
LUMO+4	(37)	-0.08781	-2.3894
LUMO+3	(36)	-0.09696	-2.6384
LUMO+2	(35)	-0.10981	-2.9881
LUMO+1	(34)	-0.13825	-3.7620
LUMO	(33)	-0.15883	-4.3220

HOMO (32) -0.43537 -11.8470
HOMO-1 (31) -0.45485 -12.3771
HOMO-2 (30) -0.52884 -14.3905
Gap (au ; eV ; nm) 0.27654 ; 7.5250 ; 164.76

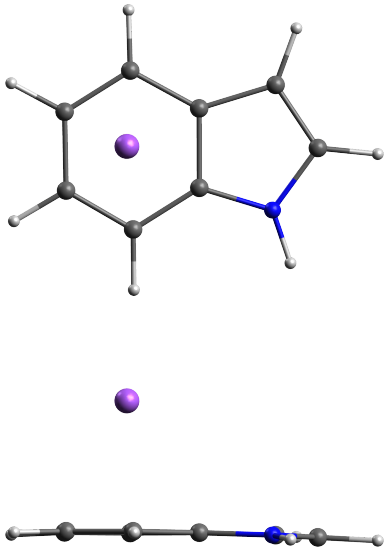
Zero-point correction= 0.132094
(Hartree/Particle)
Sum of electronic and zero-point Energies= -374.359631 au

Counterpoise fragments' energies:
BN-indole -367.160143381 au
Li^+ -7.26442813609 au
Isolated fragments' energies at complex geometry:
BN-indole -367.158387072 au
Li^+ -7.26426144960 au

10. Indole-π_6...Na^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.477827	-0.705781	-0.199824
2	6	0.454913	0.712694	-0.309237
3	6	-0.763456	1.351097	-0.599719
4	6	-1.908778	0.585066	-0.752122
5	6	-1.867322	-0.818429	-0.618427
6	6	-0.679817	-1.478459	-0.340592
7	6	2.565813	0.056007	0.062558
8	6	1.801548	1.165140	-0.128419
9	1	-0.799225	2.425611	-0.735894
10	1	-2.844668	1.063710	-1.012698
11	1	-2.771964	-1.393670	-0.771153
12	1	-0.647899	-2.559806	-0.278169
13	1	2.119608	-2.015964	0.088866
14	1	3.630263	-0.027619	0.214726
15	1	2.157229	2.181672	-0.166996
16	7	1.776082	-1.070414	0.035739
17	11	-1.250936	0.237723	1.793257



SCF Energy of neutral: -525.8192367450000 au
Dipole moment (Debye):
x= 2.194 y= 0.765 z= 6.161 Norm= 6.585

HOMO = -11.318 eV LUMO = -4.488 eV gap(HOMO-LUMO) = 6.830 eV

MO	(#)	au	eV
LUMO+4	(41)	-0.09377	-2.5516
LUMO+3	(40)	-0.09594	-2.6107
LUMO+2	(39)	-0.10091	-2.7459

```
  LUMO+1   ( 38)  -0.13406   -3.6480
  LUMO     ( 37)  -0.16493   -4.4880
  HOMO     ( 36)  -0.41594  -11.3183
  HOMO-1   ( 35)  -0.44282  -12.0497
  HOMO-2   ( 34)  -0.50541  -13.7529
  HOMO-3   ( 33)  -0.55549  -15.1157
Gap (au ; eV ; nm)        0.25101 ;    6.8303 ;    181.52

Zero-point correction=                          0.132506
(Hartree/Particle)
Sum of electronic and zero-point Energies=        -525.684741 au
```

Counterpoise fragments' energies:

Indole -363.703547338 au

Na^+ -162.064803965 au

Isolated fragments' energies at complex geometry:

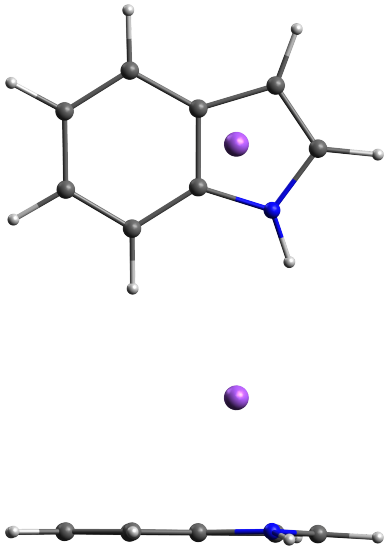
Indole -363.701990186 au

Na^+ -162.064371374 au

11. Indole-π_5...Na^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.022717	-0.792341	-0.182325
2	6	0.074988	0.575983	-0.535831
3	6	-1.113625	1.328067	-0.529477
4	6	-2.289978	0.701725	-0.182342
5	6	-2.317820	-0.661883	0.163315
6	6	-1.171286	-1.426789	0.172415
7	6	2.161264	-0.279918	-0.670186
8	6	1.454858	0.878078	-0.833904
9	1	-1.105965	2.372519	-0.817117
10	1	-3.216275	1.261714	-0.187554
11	1	-3.263717	-1.122221	0.418630
12	1	-1.201072	-2.480049	0.423260
13	1	1.566571	-2.257519	-0.197741
14	1	3.206573	-0.481159	-0.847161
15	1	1.852854	1.812540	-1.198294
16	7	1.315947	-1.281700	-0.242971
17	11	1.092972	0.720776	1.790616



SCF Energy of neutral: -525.8136993920000 au
Dipole moment (Debye):
x= -4.482 y= -0.703 z= 6.150 Norm= 7.642

HOMO = -11.394 eV LUMO = -4.654 eV gap(HOMO-LUMO) = 6.740 eV

MO	(#)	au	eV
LUMO+4	(41)	-0.08598	-2.3396
LUMO+3	(40)	-0.09728	-2.6471
LUMO+2	(39)	-0.10318	-2.8077

```
  LUMO+1   ( 38)  -0.12394   -3.3726
  LUMO     ( 37)  -0.17103   -4.6540
  HOMO     ( 36)  -0.41872  -11.3940
  HOMO-1   ( 35)  -0.43641  -11.8753
  HOMO-2   ( 34)  -0.50458  -13.7303
  HOMO-3   ( 33)  -0.54715  -14.8887
Gap (au ; eV ; nm)        0.24769 ;    6.7400 ;    183.95

Zero-point correction=                          0.132188
(Hartree/Particle)
Sum of electronic and zero-point Energies=        -525.679593 au
```

Counterpoise fragments' energies:

Indole -363.703322209 au

Na^+ -162.064809043 au

Isolated fragments' energies at complex geometry:

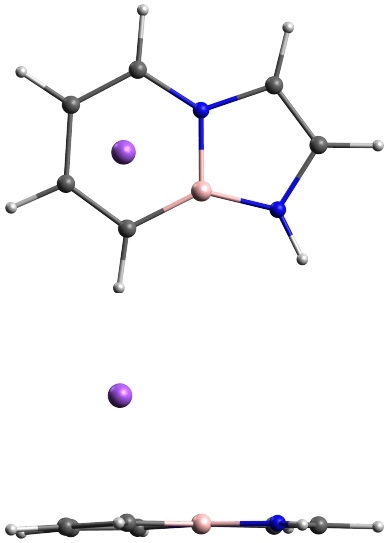
Indole -363.701840922 au

Na^+ -162.064371374 au

12. BN-indole-π_6...Na^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.616447	-1.459392	-0.358981
2	6	1.815773	-0.862318	-0.635243
3	6	1.924762	0.557951	-0.772875
4	6	0.853552	1.403513	-0.602162
5	6	-2.593498	-0.007246	0.093320
6	6	-1.811235	-1.105944	0.065248
7	1	0.502114	-2.535162	-0.292660
8	1	2.681245	-1.487779	-0.807710
9	1	2.898112	0.957741	-1.040495
10	1	1.011252	2.466711	-0.744889
11	1	-2.280638	2.043072	-0.163305
12	1	-3.660120	0.030621	0.248561
13	1	-2.081640	-2.142754	0.177143
14	7	-1.844397	1.137482	-0.114753
15	5	-0.485539	0.765103	-0.329434
16	7	-0.494267	-0.692760	-0.177549
17	11	1.353929	0.233600	1.780073



SCF Energy of neutral: -529.2751874739999 au
Dipole moment (Debye):
x= -2.593 y= -0.562 z= 6.001 Norm= 6.561

HOMO = -11.178 eV LUMO = -4.555 eV gap(HOMO-LUMO) = 6.624 eV

MO	(#)	au	eV
LUMO+4	(41)	-0.08104	-2.2052
LUMO+3	(40)	-0.09617	-2.6169
LUMO+2	(39)	-0.09665	-2.6300
LUMO+1	(38)	-0.14337	-3.9013

```
  LUMO    ( 37)  -0.16738   -4.5546
  HOMO    ( 36)  -0.41080  -11.1784
  HOMO-1  ( 35)  -0.43518  -11.8419
  HOMO-2  ( 34)  -0.52349  -14.2449
Gap (au ; eV ; nm)        0.24342 ;    6.6238 ;    187.18

Zero-point correction=                          0.130827
(Hartree/Particle)
Sum of electronic and zero-point Energies=        -529.142497 au
```
Counterpoise fragments' energies:

BN-indole -367.161431985 au

Na^+ -162.064836882 au

Isolated fragments' energies at complex geometry:

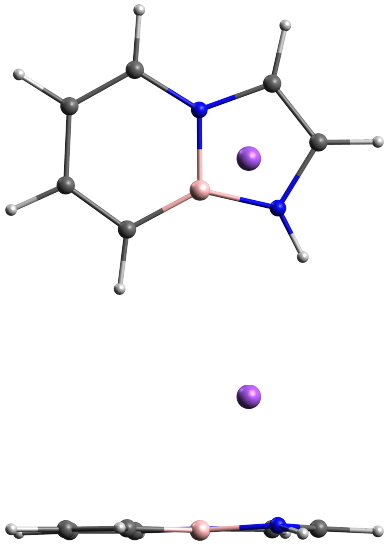
BN-indole -367.160034439 au

Na^+ -162.064371374 au

13. BN-indole-π_5...Na^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	-1.161736	1.410933	0.023739
2	6	-2.318143	0.716775	0.145684
3	6	-2.365219	-0.703079	-0.017439
4	6	-1.265718	-1.457217	-0.320224
5	6	2.123439	0.224706	-0.732802
6	6	1.290410	1.248444	-0.436021
7	1	-1.098372	2.486632	0.129621
8	1	-3.228439	1.258304	0.363253
9	1	-3.333124	-1.179435	0.099109
10	1	-1.387141	-2.526883	-0.443545
11	1	1.846436	-1.825693	-1.071838
12	1	3.173027	0.277323	-0.978058
13	1	1.502784	2.305103	-0.392424
14	7	1.424447	-0.983853	-0.711992
15	5	0.018428	-0.705064	-0.496699
16	7	-0.000797	0.729656	-0.256145
17	11	1.331722	-0.375820	1.779702



SCF Energy of neutral: -529.2701887130000 au
Dipole moment (Debye):
x= 5.204 y= -0.838 z= 5.535 Norm= 7.643

HOMO = -11.443 eV LUMO = -4.674 eV gap(HOMO-LUMO) = 6.769 eV

MO	(#)	au	eV
LUMO+4	(41)	-0.08110	-2.2068
LUMO+3	(40)	-0.09800	-2.6667
LUMO+2	(39)	-0.10463	-2.8471
LUMO+1	(38)	-0.12489	-3.3984

LUMO	(37)	-0.17176	-4.6738
HOMO	(36)	-0.42052	-11.4429
HOMO-1	(35)	-0.42975	-11.6941
HOMO-2	(34)	-0.51137	-13.9151
HOMO-3	(33)	-0.52964	-14.4122

Gap (au ; eV ; nm) 0.24876 ; 6.7691 ; 183.16

Zero-point correction= 0.130676
(Hartree/Particle)
Sum of electronic and zero-point Energies= -529.137485 au

Counterpoise fragments' energies:

BN-indole	-367.160562630 au
Na^+	-162.064858223 au

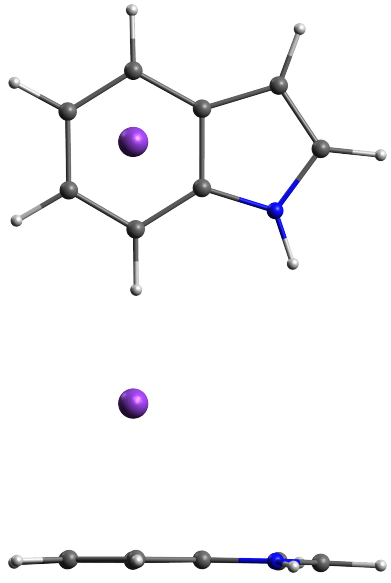
Isolated fragments' energies at complex geometry:

BN-indole	-367.159022309 au
Na^+	-162.064371374 au

14. Indole-π_6...K^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	-0.728894	0.538380	0.573102
2	6	-0.676321	0.197607	-0.804123
3	6	0.460980	0.552095	-1.546996
4	6	1.505427	1.201170	-0.912326
5	6	1.440515	1.512423	0.459763
6	6	0.327694	1.187823	1.217072
7	6	-2.658985	-0.453441	0.019386
8	6	-1.919422	-0.440656	-1.122606
9	1	0.505264	0.348539	-2.610695
10	1	2.371637	1.511759	-1.483760
11	1	2.259328	2.050118	0.921757
12	1	0.265817	1.467303	2.262324
13	1	-2.308993	0.301090	1.968534
14	1	-3.657783	-0.823611	0.188380
15	1	-2.233951	-0.809954	-2.085076
16	7	-1.944750	0.118301	1.047568
17	19	1.573998	-1.612935	0.326628



SCF Energy of neutral: -963.5099480540000 au
Dipole moment (Debye):
x= 3.691 y= 5.832 z= -2.758 Norm= 7.433

HOMO = -10.996 eV LUMO = -3.953 eV gap(HOMO-LUMO) = 7.043 eV

MO	(#)	au	eV
LUMO+4	(45)	-0.08548	-2.3260
LUMO+3	(44)	-0.09168	-2.4947
LUMO+2	(43)	-0.09554	-2.5998

```
  LUMO+1   ( 42)  -0.12093   -3.2907
  LUMO     ( 41)  -0.14527   -3.9530
  HOMO     ( 40)  -0.40410  -10.9961
  HOMO-1   ( 39)  -0.42858  -11.6623
  HOMO-2   ( 38)  -0.49132  -13.3695
  HOMO-3   ( 37)  -0.54267  -14.7668
Gap (au ; eV ; nm)        0.25883 ;    7.0431 ;    176.04

Zero-point correction=                          0.132191
(Hartree/Particle)
Sum of electronic and zero-point Energies=        -963.376498 au
```

Counterpoise fragments' energies:

Indole -363.703289471 au

K^+ -599.769520006 au

Isolated fragments' energies at complex geometry:

Indole -363.702126998 au

K^+ -599.769423689 au

15. Indole-π_5…K^+

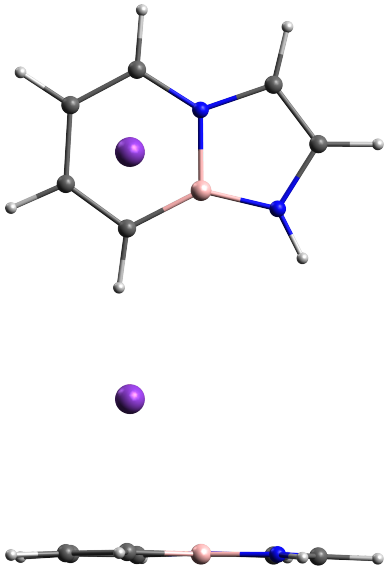
Cation, ground state, CAM-B3LYP/6-311G(d,p)

The optimization process leads to the previous indole-π_6…K^+ structure.

16. BN-indole-π_6...K^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

Center Number	Atomic Number	Coordinates (Angstroms) X	Y	Z
1	6	0.314823	0.973659	-1.335211
2	6	1.427311	1.406006	-0.672477
3	6	1.520744	1.329583	0.752374
4	6	0.524891	0.800410	1.533396
5	6	-2.683477	-0.440807	-0.056015
6	6	-1.944727	-0.050177	-1.115129
7	1	0.201973	1.063510	-2.409545
8	1	2.224645	1.867332	-1.239772
9	1	2.417085	1.734853	1.212233
10	1	0.658995	0.803272	2.609479
11	1	-2.412671	-0.443846	2.016467
12	1	-3.683730	-0.843888	-0.069846
13	1	-2.198436	-0.046201	-2.162281
14	7	-1.986686	-0.251986	1.125442
15	5	-0.720338	0.335339	0.824579
16	7	-0.716715	0.425103	-0.635019
17	19	1.597908	-1.638716	-0.113380



SCF Energy of neutral: -966.9661946990000 au
Dipole moment (Debye):
x= -3.864 y= 6.233 z= -0.607 Norm= 7.359

HOMO = -10.908 eV LUMO = -3.999 eV gap(HOMO-LUMO) = 6.910 eV

MO	(#)	au	eV
LUMO+4	(45)	-0.07930	-2.1579
LUMO+3	(44)	-0.09244	-2.5154
LUMO+2	(43)	-0.09433	-2.5669

```
  LUMO+1   ( 42)  -0.12867   -3.5013
  LUMO     ( 41)  -0.14695   -3.9987
  HOMO     ( 40)  -0.40087  -10.9082
  HOMO-1   ( 39)  -0.42060  -11.4451
  HOMO-2   ( 38)  -0.50801  -13.8237
  HOMO-3   ( 37)  -0.52826  -14.3747
Gap (au ; eV ; nm)        0.25392 ;    6.9095 ;    179.44

Zero-point correction=                          0.130549
(Hartree/Particle)
Sum of electronic and zero-point Energies=        -966.834457 AU
```

Counterpoise fragments' energies:

BN-indole $\quad$ -367.161373848 au

K^+ $\quad$ -599.769533975 au

Isolated fragments' energies at complex geometry:

BN-indole $\quad$ -367.160295120 au

K^+ $\quad$ -599.769423689 au

17. BN-indole-π_5…K^+

Cation, ground state, CAM-B3LYP/6-311G(d,p)

The optimization process leads to the previous BN-indole-π_6…K^+ structure.

18. Binding energy calculations

According to:
Kirschner, Karl N., Jennifer B. Sorensen, et J. Phillip Bowen. « Calculating Interaction Energies Using First Principle Theories: Consideration of Basis Set Superposition Error and Fragment Relaxation ». *Journal of Chemical Education* 84, nº 7 (juillet 2007): 1225. https://doi.org/10.1021/ed084p1225.

ΔE_{Rig} corresponds to the binding energy taking into account the basis set superposition error (BSSE).

Calculation method: CAM-B3LYP/6-311G(d,p)

	ΔE (kcal/mol)									
A...B	$\Delta E(AB)$	$\Delta E_{SP}(AB)$	$\Delta E_{fCP}(AB)$	**$\Delta E_{Rig}(AB)$**	$\Delta E_{BSSE}(A)$	ΔE_{BSSE} (B)	$\Delta E_{BSSE}(AB)$	$\Delta E_{Rel}(A)$	$\Delta E_{Rel}(B)$	$\Delta E_{Rel}(AB)$
Indole-π_6...Li^+	-49.334	-49.800	-48.740	**-48.273**	0.963	0.098	1.061	0.466	0.000	0.466
BN-indole-π_6...Li^+	-46.798	-47.711	-46.690	**-45.777**	0.913	0.108	1.021	0.913	0.000	0.913
Indole-π_5...Li^+	-44.132	-45.049	-43.784	**-42.867**	1.170	0.096	1.265	0.917	0.000	0.917
BN-indole-π_5...Li+	-42.875	-44.553	-43.346	**-41.668**	1.102	0.105	1.207	1.678	0.000	1.678
Indole-π_6...Na^+	-32.749	-33.180	-31.931	**-31.500**	0.977	0.271	1.249	0.431	0.000	0.431
BN-indole-π_6...Na^++	-31.222	-31.866	-30.697	**-30.053**	0.877	0.292	1.169	0.644	0.000	0.644
Indole-π_5...Na^+	-29.274	-29.799	-28.594	**-28.070**	0.930	0.275	1.204	0.525	0.000	0.525
BN-indole-π_5...Na^+	-28.085	-29.364	-28.092	**-26.813**	0.967	0.306	1.272	1.279	0.000	1.279
Indole-π_6..K^+	-23.749	-24.095	-23.305	**-22.959**	0.729	0.060	0.790	0.345	0.000	0.345
BN-indole-π_6..K^+	-22.408	-22.889	-22.143	**-21.662**	0.677	0.069	0.746	0.481	0.000	0.481
Indole-π_5...K^+	—	—	—	**—**	—	—	—	—	—	—
BN-indole-π_5...K^+	—	—	—	**—**	—	—	—	—	—	—